

Savoir, pour prévoir, afin de pouvoir.  
Auguste Comte

Prévoir l'avenir, c'est le déterminer.  
Auteur inconnu

---

## **Systeme d'information décisionnel : à quoi cela sert-il ?**

---

Dans l'entreprise le système d'information (SI) a pour objectif de faciliter l'établissement et la mise en œuvre de la stratégie, en particulier de concrètement supporter la réalisation des activités. Il est construit à partir des exigences des métiers, des processus définis par l'entreprise, et il est constitué de l'ensemble des moyens (humains, logiciels, matériels) utilisés pour collecter, stocker, traiter et communiquer les informations.

Il est d'usage de distinguer trois types différents de SI, les systèmes supportant la conception des produits (calcul numérique, CAO, ...), les systèmes industriels (conduite de machines, contrôle de process industriel, ...) et les systèmes de gestion. Ces derniers couvrent toutes les activités de gestion du fonctionnement de l'entreprise (marketing, vente, achat, production, logistique, finance, ressources humaines, R&D). Pour des raisons techniques, qui existent toujours en partie aujourd'hui, les systèmes d'information de gestion ont été historiquement structurés en deux sous systèmes : l'un dit opérationnel qui prend en charge la réalisation des opérations au jour le jour, et l'autre dit décisionnel qui fournit des informations pour définir la stratégie, piloter les opérations et analyser les résultats.

Un système décisionnel est donc avant tout un moyen qui a pour but de faciliter la définition et la mise en œuvre de stratégies gagnantes. Mais il ne s'agit pas de définir une stratégie une fois pour toute, mais d'être à même de continuellement s'adapter à son environnement, et de le faire plus vite que ses concurrents. Pour cela il convient de bien comprendre son environnement, d'ajuster ses interactions avec lui en faisant les meilleurs choix de cibles et d'actions. Concrètement le chemin à suivre peut être caractérisé par les quatre objectifs

suivants : comprendre son environnement, se focaliser sur des cibles, aligner son organisation et mettre en œuvre les plans d'actions nécessaires.

Un système décisionnel va en particulier aider au pilotage des plans d'actions (prévision, planification, suivi), à l'apprentissage (acquisition de savoir faire, de connaissances, de compétences) et à la réalisation d'innovations incrémentales (adaptation du modèle d'affaires : produits/services, organisation, etc. ...). Les systèmes décisionnels traditionnels permettent de faire l'analyse des activités déjà réalisées et d'en tirer des enseignements pour les activités futures, pour cela ils utilisent des données plus ou moins récentes (au mieux mises à jour quotidiennement). Les systèmes décisionnels plus avancés gèrent des données plus fraîches (certaines sont mises à jour en quasi temps réel), automatisent des décisions et supportent en temps réel des opérations (centre d'appels, web par exemple).

## **Du pilotage stratégique à l'intelligence économique**

Le pilotage stratégique beaucoup de monde en parle, mais peu d'entreprises savent réellement en tirer parti. Elles n'y voient trop souvent qu'une version améliorée des bons vieux tableaux de bord. Ce faisant elles oublient l'apport majeur du pilotage stratégique qui est de mobiliser l'énergie collective dans le cadre d'un processus d'apprentissage continu de son environnement.

Mais comment parler d'énergie collective quand on voit telle entreprise vanter les mérites de son nouveau système d'information, reposant sur 17 objectifs stratégiques et une quarantaine d'indicateurs, qui sont analysés par les cinquante cadres dirigeants gravitant autour de la direction générale. Pourquoi cinquante personnes seulement, alors que cette vision et ces objectifs stratégiques devraient être partagés par l'ensemble de l'entreprise ?

L'expérience montre que si les entreprises sont aujourd'hui en mesure de définir leur vision et leurs objectifs stratégiques, elles ont en revanche toujours autant de mal à les mettre en œuvre et les diffuser dans l'ensemble de leurs équipes. Au mieux parvient-on à mobiliser la direction générale, le marketing et les forces commerciales, par nature plus sensibles à leur environnement. Mais par exemple les équipes de production sont encore trop souvent mises à l'écart. Pourquoi ? Parce que les entreprises sont toujours très cloisonnées. Parce que chaque service interprète les objectifs stratégiques en fonction de sa culture et de ses propres intérêts. Parce que l'empilement des structures et des systèmes d'informations complique la mobilisation de l'ensemble de l'effectif, autour d'objectifs globaux.

Or, les nouveaux outils de BI facilitent la mise au point d'indicateurs fédérant toutes les couches de l'entreprise. La vision stratégique de la direction générale doit en effet être diffusée et déclinée dans tous les services de l'entreprise : il s'agit de fixer à chacun des objectifs qui lui « parlent ». Parce qu'un chef d'équipe sur une chaîne de montage se sentira infiniment plus concerné par un objectif de réduction des délais de mise en route de la chaîne ou par un ratio de défauts, que par un objectif d'accroissement de part de marché au plan mondial.

Le véritable enjeu du pilotage stratégique consiste donc à décliner une vision globale en une batterie de tableaux de bord pertinents : chaque tableau de bord étant un affinage de l'objectif général adapté aux réalités de chaque équipe. La mise au point de ces tableaux de bord repose sur un système décisionnel performant, permettant d'exploiter conjointement l'ensemble des informations internes à l'entreprise, des informations issues des veilles technologique, commerciale et sociétale, ainsi que du benchmarking, de façon à mettre au point des tableaux de bord prospectifs et véritablement orientés vers l'amélioration de la performance sur un marché donné.

Pour être efficaces, ces outils de gestion de la performance doivent concerner le plus de gens possible. Par exemple le système de pilotage stratégique du Réseau de Distribution d'un Grand Constructeur Automobile mobilise ainsi plus de 2 500 salariés autour d'une batterie de 2 000 indicateurs. Dans une grande entreprise, un bon pilotage stratégique doit fédérer plusieurs milliers de collaborateurs autour d'autant d'indicateurs (à raison d'une dizaine d'indicateurs au maximum par équipe, et d'une cinquantaine pour la direction générale).

Dans ces conditions et s'il est porté par une direction générale et des contrôleurs de gestion motivés, le pilotage stratégique peut devenir un thème aussi porteur et novateur pour les entreprises que l'ont été ceux de la qualité totale et de la gestion des relations client.

## **Besoins et tendances en matière de décisionnel dans un contexte de globalisation**

Le marché du décisionnel continue à se porter à merveille. La raison fondamentale du maintien d'une haute demande tient à la capacité de la BI à répondre aux besoins des directions générales d'aujourd'hui.

Dans les grandes entreprises les directions générales poursuivent principalement trois objectifs. Premièrement globaliser les activités, c'est-à-dire concrètement définir une stratégie globale, assurer l'unité de sa mise en œuvre par toute l'entreprise qui doit manœuvrer comme un seul homme, et piloter l'ensemble au plus près sans délai. Deuxièmement optimiser les processus, c'est-à-dire se focaliser sur les éléments clés du métier de l'entreprise (création de l'offre, production, commercialisation et gestion des ressources), avec concrètement la fin des approches par fonction au profit d'une intégration des visions (commerciales, financières, ...), pour optimiser la valeur ajoutée prévue dans le cadre de la stratégie. Troisièmement rationaliser les moyens, c'est-à-dire les dimensionner, les localiser en fonction des objectifs stratégiques, avec concrètement un développement du travail en réseaux mondiaux et la fin des baronnies, des potentats locaux et des degrés de liberté des filiales.

Plus précisément si l'on prend le domaine financier d'une entreprise il s'agit de maîtriser et d'harmoniser les pratiques à travers toutes les filiales (par exemple la société NCR est présente dans 130 pays), d'accélérer la publication des résultats (souvent actuellement il faut de 5 à 8 jours ; la nouvelle cible est de 2 jours), de réduire les coûts de la fonction financière (actuellement souvent > 1% du chiffre d'affaires alors que la nouvelle cible est d'environ 0,5%), de réduire les encours clients, de réduire les stocks, etc. Dans les entreprises qui globalisent leurs activités, tous les domaines Marketing/Ventes, Production, Gestion des Ressources, R&D font l'objet des mêmes approches de redéfinition des objectifs, des organisations et des moyens.

En terme de système décisionnel pour répondre aux besoins actuels des grandes entreprises décrits ci-dessus, il convient comme il a été vu au chapitre précédent d'être à même de : 1°) équiper de plus en plus d'utilisateurs différents (Directions, Experts en analyse de données, Analystes d'une fonction, Responsables opérationnels, Opérationnels, ...). 2°) intégrer les données dans un entrepôt de données d'entreprise (EDW) : intégrer les ODS, consolider les Data Marts, assurer via l'EDW la qualité des données. 3°) traiter de plus en plus de données : taille des disques, puissance des systèmes, ETL & ELT, EDW et mise en place de Data Marts logiques. 4°) intégrer du support décisionnel dans la gestion des opérations (rendre l'entrepôt de données actif – ADW) : multiplication des scorings, développement de services internet (SOA). 5°) intégrer dans une architecture d'entreprise les différents outils de la BI (Oracle, IBM, Microsoft, SAP, SAS, ...). 6°) développer de nouvelles applications de gestion du client (Calcul de la valeur client, compte d'exploitation Client en BtoC, ...), de gestion des produits (par exemple problématique des garanties : maîtrise aval et amont de la vie des produits), etc.

Concrètement pour les grandes entreprises cela veut dire que leur système décisionnel doit : être disponible 24h sur 24, 7 jours sur 7, 52 semaines par an ; alimenter en données de qualité des milliers d'utilisateurs (actuellement ~12 % des effectifs d'une grande entreprise sont dotés de moyens d'aide à la décision) ; stocker souvent des dizaines de téras de données, mises à jour pour certaines d'entre elles en quasi temps réel par rapport à l'opérationnel, intégrées dans un gisement fondé sur un modèle normalisé couplé avec de multiples modèles en étoiles ; répondre aux besoins de l'ensemble des domaines fonctionnels de l'entreprise, en autorisant simultanément les activités de reporting, des requêtes ad hoc, de l'olap, de l'analyse de données qui manipulent des téras de données utiles et autorisent des vues logiques complexes (>15 jointures).

La valeur d'un système décisionnel pour une entreprise est déterminée par sa capacité à aider à être plus agile et pour cela dans de nombreux domaines de l'aide à la décision la vitesse est primordiale : vitesse d'acquisition, d'analyse et d'action. A ce jeu il s'agit de réduire le plus possible les latences entre l'occurrence d'un événement métier, la capture des données correspondantes, la mise en cohérence et en perspective, la production d'une information et l'action. L'expérience montre que seule une architecture de système décisionnel fondée sur un gisement centralisé de données détaillées d'entreprise peut permettre de répondre à ce défi.

Au jeu de la globalisation, une entreprise équipée d'une bonne solution décisionnelle a un avantage certain et pour cela son système doit être architecturé de façon à garantir les performances à court terme, l'évolutivité, la pérennité des investissements, une maîtrise des coûts et des risques.

### **Etre dans le vent, c'est une ambition de feuille morte : parlez métier**

En tant que responsable Marketing je suis bien placé pour savoir que le vocabulaire suit des modes et vieillit vite. Infocentre, EIS, SIAD, CPM, Scorecarding, sont autant de mots qui se périment et ne permettent plus de correctement décrire les pratiques et les projets des entreprises d'aujourd'hui. L'utilisation de données historiques se développe dans l'entreprise en surfant sur la loi de Moore. Il y a quelques années la presse publiait régulièrement des listes d'entreprises qui avaient un entrepôt de données de plus d'un Tétra, aujourd'hui certaines entreprises ont déjà des entrepôts de données de plus d'un Péta.

Ce développement effréné vient du fait que les entreprises ont découvert que la première condition de la conduite intelligente des affaires, c'est d'avoir de la mémoire, que les décisions essentielles pour la vie des organisations se prennent à tous les niveaux des structures, et que l'information à d'autant plus de pertinence qu'elle colle à l'événement métier et qu'il convient donc de la capturer, de la mettre en cohérence, en perspective et de la délivrer à la vitesse grand V. Exit donc tous les systèmes au rythme batch et aux données agrégées, tous ses data marts optimisés qui sont obsolètes le jour de leur recette.

La problématique d'aujourd'hui se formule de la façon suivante : comment mettre à disposition des opérationnels, des planificateurs, des managers et des stratèges des différentes fonctions de l'entreprise des informations de qualité qui concernent leur activité, en répondant à leurs besoins de volume de données à stocker, de fraîcheur, de possibilité de traitement (production d'indicateurs, réalisation d'analyses ad hoc et de simulation, ...) et de disponibilité. Face à ces besoins, les entreprises ont généralement multipliés les systèmes et

mis en place des moyens spécifiques, disparates et souvent incohérents pour faciliter la détermination et le pilotage des stratégies ; pour définir les objectifs, les moyens et les budgets ; pour planifier et suivre les opérations (marketing, commercial, production, ...) ; pour s'aider dans l'exécution des tâches opérationnelles et notamment dans les interactions avec les partenaires de l'entreprise en particulier les clients et les fournisseurs.

Les bons objectifs à se fixer sont en premier de créer une vue consistante des données à travers l'entreprise, puis de mettre en place des moyens qui permettent d'obtenir plus rapidement des réponses aux questions métiers. Pour cela il convient de construire un entrepôt de données dynamique qui enregistre les événements métier au fur et à mesure qu'ils sont émis par les applications opérationnelles, et qui les met à disposition pour supporter les décisions opérationnelles, de pilotage, de management ou de stratégie.

Si nous prenons le processus de la relation avec les clients, le volume des événements à gérer pour une grande banque se chiffre en centaines de millions d'interactions par an, avec par exemple pour une grande banque française : 100 millions de contacts en agence, 100 millions de sessions aux distributeurs ou guichets automatiques, 40 millions de visites internet, 20 millions d'accès au serveur vocal, 5 millions de contacts au centre d'appels. Une bonne gestion de ces événements et de leur historique doit permettre une plus grande efficacité de la force de vente, du marketing en particulier par une meilleure connaissance des clients et une interaction multi-canal plus cohérente.

Pour en revenir au vocabulaire qui vieillit et pour éviter dans l'avenir de nous tromper de vocabulaire, je conseille de ne pas suivre la mode de la BI ou même d'utiliser les mots de la BI. Il convient de parler de données, d'applications métier et concrètement par exemple dans le secteur de la distribution : d'assortiment, de stock, de fournisseur ou encore de façon plus fine de la suppression d'un article, des affinités produits, des ventes perdues toutes choses qui se traquent à travers les événements que les systèmes opérationnels gèrent.

### Exemple d'approche métier du décisionnel

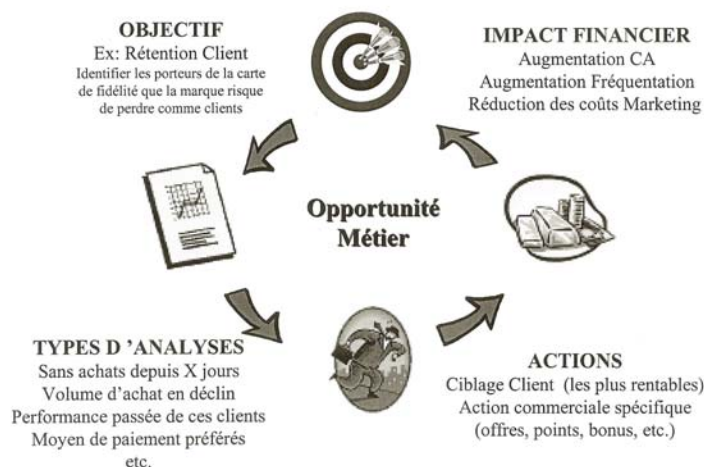


Illustration – 7

## **L'entrepôt de données un des outils de l'intelligence économique**

Maîtriser l'information est plus que jamais un moteur de compétitivité pour l'entreprise, mais aussi pour les Etats. Dans le cadre de la mondialisation et du passage d'une logique d'optimisation de la production à une logique d'innovation, la puissance d'une entreprise ou d'un Etat est liée à la maîtrise de toute la chaîne d'information, pour les trois raisons principales décrites ci-dessous.

Les modes de consommation ont évolué. Nous sommes passés d'un marché de masse à une production diversifiée qui cherche à répondre avec finesse à la demande et ce dans toutes les branches d'activité (industrie, assurance, etc.). Aujourd'hui le client n'est plus face à l'entreprise, il pénètre dans l'entreprise. Autrefois, la clientèle était disponible, en position d'attente, la concurrence étant limitée, voire nulle. Aujourd'hui on est à l'écoute du client dans une posture proactive.

Le marché est devenu concurrentiel après des années de cloisonnement par l'Etat nation, particulièrement en France, pays dont 70 % des échanges extérieurs concernaient en 1970 uniquement son vieil empire colonial. Aujourd'hui, près de 70 % des échanges extérieurs français ont lieu au sein de l'Europe. Globalement nous sommes passés en 40 ans d'un marché protégé à un marché ouvert.

Le progrès technique s'accélère et justifie une meilleure maîtrise de l'information. Les investissements immatériels (logiciels, brevets, etc.) ont largement pris le pas sur les investissements matériels. Ils constituent un facteur essentiel du développement de la concurrence.

Cette exigence de maîtrise de l'information, nous amène à développer l'approche de l'Intelligence Economique. Il s'agit à la fois d'un mode de pensée – l'information doit être créée et valorisée – et d'un mode d'action – l'information doit être partagée et utilisée à bon escient. La pratique de l'intelligence économique s'appuie sur quatre idées fondamentales.

La maîtrise des connaissances : connaître et protéger. L'identité d'une entreprise tient plus à son savoir et à son savoir-faire qu'à son produit. Souvent, elle est potentiellement capable de faire autre chose, en basant sa compétitivité davantage sur ce qu'elle sait faire que sur ce qu'elle fait déjà. Enrichir le système en place n'est pas simple et il convient de faire une distinction entre protection de l'information stratégique et culture du secret. Dans l'entreprise, beaucoup trop d'informations sont classées "top secret". Il est important de hiérarchiser le niveau de confidentialité afin de ne pas étouffer la communication indispensable au développement de l'entreprise, voire à sa survie.

La compréhension des opportunités, des risques et des menaces. La fonction marketing et la veille permettent d'identifier des opportunités de marché, mais aussi des « opportunités capitalistiques », surtout en période de concentration. Le risque peut provenir de l'obsolescence des savoir-faire ou d'une mauvaise gestion des compétences. Les menaces – la corruption, le piratage, la copie, l'espionnage et toute tentative de déstabilisation – ne sont fatales que si le concurrent peut identifier des objectifs clairs.

La coordination des acteurs et des activités. Cette coordination est réalisée en interne, en sensibilisant le personnel aux orientations de l'entreprise, mais aussi en externe, par la constitution d'un réseau de fournisseurs, de clients, de centres de recherche, etc. L'entreprise

n'existe pas seule. 50 % des entreprises innovantes, avec un taux de croissance annuel supérieur à 15 %, ont au moins trois vrais partenaires stratégiques.

La stratégie d'influence. Il s'agit de la capacité à maîtriser son environnement. Et cela n'est pas uniquement à la portée des grandes entreprises : même une petite structure peut influencer les directives européennes, via la représentation locale. Si un maire ne souhaite pas qu'une usine de cinquante personnes ferme ses portes, il en parle au conseiller régional, qui en parle au député, qui en parle au ministre, etc. Mais le mieux reste encore d'anticiper, de repérer les réseaux d'influence, les alliés potentiels. Tous les salariés sont évidemment concernés.

Dans sa dimension « Information », l'Intelligence Economique s'appuie sur des techniques de veille (technologique, commerciale, ...), de renseignement, de benchmarking, de gestion des performances et des connaissances. La veille technologique, qui surveille le marché pour en détecter les signaux faibles et les signaux forts d'évolution. Le benchmarking, qui consiste à voir et comparer les solutions des divers concurrents. Le renseignement, c'est-à-dire la capacité à faire émerger l'information confidentielle (appel d'offres, nouvelle technologie, etc.). Le knowledge management, à savoir se connaître soi-même, identifier l'information à maîtriser, celle à partager et celle à protéger.

Dans sa dimension « Action », l'Intelligence Economique s'appuie sur des techniques d'influence, de lobbying, de communication, de recrutement, de gestion des compétences, des informations et de la confidentialité.

D'un point de vue technologique, l'Intelligence Economique utilise les moyens d'information et de communication d'aujourd'hui : gestion de documents, data & text mining, moteur de recherche, business intelligence, intra/extra/internet et entrepôt de données.

Tous ces outils sont excellents lorsque l'information se met au service de la stratégie et il s'agit donc en particulier de savoir intégrer des données en provenance de sources très variées :

- Les fournisseurs d'information : Presse, librairies, bibliothèques, banques de données, centres d'information,
- Les nouveaux réseaux : Internet, ...
- Les institutions : Administration, organisations internationales, associations, centres de recherche, instituts, fondations, universités,
- Les groupes partenaires : Sous-traitants, fournisseurs, clients, concurrents, groupements professionnels, financiers,
- Les consultants et experts : Ecoles d'ingénieurs, universitaires, experts, centres techniques, consultants, bureaux d'études,
- Les manifestations : Colloques, salons, voyages d'études,
- Les réseaux personnels : Clubs, collègues, confrères,
- Les sources fortuites : Train, avion, la rue, les stagiaires, ...
- Les sources internes : ... encore faut-il ne pas les oublier.

Mais attention à la légalité de vos approches.

En conclusion, le moyen le plus sûr de conforter sa compétitivité au niveau mondial est certainement de transformer son organisation, en système d'intelligence économique, en



associant largement ses employés, ses partenaires sous la bannière du chef d'entreprise, et en mettant en place les moyens technologiques nécessaires pour accueillir et traiter l'information, c'est-à-dire un entrepôt de données d'entreprise.

### **La BI en trois étapes de bon sens : partez lentement pour aller vite**

Alors que les entreprises gèrent d'immenses quantités de données, les experts informatiques ne cessent de chanter les louanges du concept de la Business Intelligence (BI). Avoir un accès rapide à une compréhension toujours plus pointue des clients, des opérations et des dangers qui menacent, est effectivement une perspective extrêmement séduisante. Sous la pression des demandes constantes de réduction des coûts, de diminution des risques, de prise en compte des exigences légales et d'amélioration des bénéfices, les responsables métier sont particulièrement attirés par les possibilités offertes par la BI.

Le défi est de se rendre d'un point à un autre et les nombreuses pierres d'achoppement potentielles peuvent rendre le voyage particulièrement difficile. Certains ont échoué en essayant d'en faire trop d'un coup, en voulant que toutes les données de leur entreprise soient disponibles immédiatement pour un large éventail d'interrogations potentielles. Cela représente un investissement énorme en temps, en ressources et ces entreprises doivent se battre pour simultanément absorber les coûts, mettre en place et produire une valeur. D'autres se sont montrés trop prudents avec des investissements insuffisants et des progrès trop lents. Au final une valeur ajoutée trop faible, de sorte que souvent certains utilisateurs finissent par penser qu'ils peuvent obtenir de meilleurs résultats en entretenant leur propre stock « privé » de données pour résoudre leurs problèmes spécifiques, et alors il n'y a pratiquement pas d'avantages pour l'ensemble de l'organisation à tirer d'un tel environnement de BI. Une autre erreur fréquente des organisations informatiques est de se concentrer sur une réduction forcée du volume des données, plutôt que sur le développement d'une compréhension détaillée de la façon dont les utilisateurs vont pouvoir exploiter les données et produire une véritable plus-value commerciale. Ici aussi les utilisateurs métier déçus finissent par dépendre plus que jamais de solutions de BI artisanales et disparates pour faire face à l'évolution de leurs propres besoins. Enfin dans tout les cas, quand les utilisateurs n'apprécient pas leur travail, les informaticiens se sentent dépréciés et sous-employés.

Evidemment ces exemples ne sont pas les scénarios que de nombreux promoteurs de la BI d'entreprise ont en tête lorsqu'ils évoquent leur vision. Cependant quelques entreprises ont réussi à minutieusement planifier pour leur domaine BI un parcours de bon sens et à déployer un environnement puissant et pertinent. Les meilleures pratiques de construction d'un environnement de BI ont vu le jour tant à partir de réussites que d'implémentations imparfaites.

Deux principes essentiels

1°) Il faut établir un dialogue fort et continu avec un langage commun entre les métiers et les informaticiens. Bien trop souvent les utilisateurs et les informaticiens qui construisent l'environnement de BI vivent dans des mondes totalement différents, à peine conscients des besoins ou des processus de l'autre. Soit il y a trop peu de canaux de communication, soit ils ne parlent tout simplement pas la même « langue » et se débattent pour faire comprendre leurs besoins à l'autre.

2°) La création d'un environnement de BI doit se faire étape par étape. Pour éviter d'en faire trop ou trop peu, les entreprises doivent d'abord se concentrer sur les projets prioritaires, dont la réussite peut faciliter l'adhésion des utilisateurs et ouvrir les appétits pour les projets suivants. En principe, ces premiers projets doivent répondre à l'un des deux critères suivants : permettre d'obtenir une rentabilité rapide des investissements ou avoir une signification « politique » importante pour l'organisation. Quoi qu'il en soit la démarche sera la même : capitaliser sur une réussite.

Trois étapes pour garantir un niveau d'adoption optimal et un maximum de valeur ajoutée

Première étape : La direction en collaboration avec l'informatique développe des hypothèses sur les meilleures possibilités de production de valeur ajoutée au sein de l'organisation, dans lesquelles la BI peut jouer un rôle. L'idée est que le processus de construction de la BI doit être ancré sur la question stratégique la plus urgente pour l'organisation. Pour une Banque par exemple, cela peut vouloir dire réduire de 25% le taux de clients perdus, ou développer de 10% les apports de fonds propres sur les prêts immobiliers et réduire les hypothèques, les taux de défaut de remboursement des prêts au logement de 10%. Une fois que la direction a identifié les moteurs commerciaux, elle n'a aucun mal à identifier les utilisateurs clés : des responsables de marketing et des analystes, aux représentants des services client et autres gestionnaires de relations.

Seconde étape : Clarifier et développer les hypothèses en organisant des séances de travail avec les utilisateurs du domaine. Dans cette étape, la direction et l'informatique organisent une série de réunions avec les utilisateurs concernés afin de tester et d'affiner les hypothèses originelles. Les réunions se concentrent d'abord sur la compréhension des buts et objectifs des utilisateurs. Par exemple les utilisateurs peuvent dire que dans le but d'améliorer les activités de prêts immobiliers, ils ont besoin d'avoir un profil de tous les détenteurs de prêts actuels qui ne sont pas en défaut de paiement, mais qui ont également d'autres produits de dépôt et de prêt. Il peut s'agir d'un problème de non règlement de carte de crédit ou de découvert sur un compte chèque sur lequel ils viennent d'interrompre leur ordre de virement automatique par exemple, et qui indiquerait un défaut potentiel de règlement des mensualités de leur crédit immobilier. A partir de là les utilisateurs et les informaticiens peuvent formuler des questions commerciales clés. Puis explorer de quelles données l'organisation dispose-t-elle déjà, et de quelles données a-t-elle besoin pour répondre à ces questions commerciales ? Dans quel délai l'information doit-elle être disponible pour être efficace ? Des applications sont-elles en place pour que ce soit réalisable ? Ce processus permet aux deux groupes de comprendre l'information existante, les carences des applications et de décider s'ils disposent de moyens suffisants pour réduire ces carences. Cela permet également aux deux groupes de comprendre leurs difficultés et leurs capacités réciproques, de polariser le dialogue sur l'usage final et la valeur de l'information. Enfin les deux groupes doivent considérer ces réunions comme une plateforme à partir de laquelle ils peuvent étendre leur première idée à d'autres opportunités d'amélioration. Les informaticiens peuvent dire par exemple qu'ils peuvent également fournir un historique client plus détaillé et des informations sur le profil des clients dans un délai approchant le temps réel. La question alors est, cela représente-t-il une valeur ajoutée sur la façon dont le groupe Prêts immobiliers utilise l'information sur un dépôt automatique interrompu, par exemple ? Est-ce que le fait qu'un client n'a jamais enregistré de défaut de paiement sur sa carte de crédit peut être utile ? Grâce à ce dialogue les relations prennent de la valeur, l'informatique et le commercial s'engagent dans un processus conjoint avec un objectif partagé. Ce processus et

la reconnaissance des problèmes de l'autre contributeur à réduire le « fossé » qui sépare les deux parties.

Troisième étape : construire et tester le modèle. Une fois que les besoins et les capacités effectives sont clairs, que le projet est défini, les informaticiens doivent s'appuyer sur les lignes directrices métier pour construire l'environnement BI pour ce projet particulier. La première tâche consiste à relier les questions métier aux données requises et aux applications sources des données. Les sociétés ont intérêt à organiser leurs données au sein d'un modèle logique approprié calqué sur le mode de fonctionnement du secteur concerné, pas uniquement sur la structure organisationnelle actuelle. Ceci permet de faciliter la transformation des données en information commerciale utile. A partir de là le système peut faire le lien entre des questions métier (par exemple : quel détenteur de prêt immobilier a interrompu un dépôt automatique ?), et tout un éventail d'opportunités d'améliorations métier spécifiques (réduction des défauts de paiement, amélioration de la fidélisation des détenteurs de dépôts de fonds et/ou de prêt de grande valeur, augmentation des ventes croisées sur valeur hypothécaire), qui sont liées aux buts et objectifs originaux que la direction avait définis (développer de 10% l'unité d'apport de fonds propres sur les prêts immobiliers et réduire les hypothèques et les taux de défaut de remboursement des prêts au logement du même montant). Quand cela est approprié les sociétés peuvent modéliser l'impact commercial anticipé de cette démarche, en analysant la combinaison de croissance des revenus, la réduction des coûts et/ou la diminution des risques susceptibles de se produire. Ceci contribue à déterminer les priorités des opportunités commerciales garantissant ainsi un déploiement plus efficace des ressources.

Cependant la Banque doit décider si la construction d'un tel modèle ne risque pas de ralentir le projet dans son ensemble. Au lieu d'un processus de modélisation certaines banques peuvent préférer lancer des projets pilotes préliminaires sur des marchés sélectionnés et tester les résultats obtenus avant de développer cette idée à l'échelle globale. La focalisation combinée de l'informatique et des métiers sur un modèle définitif de données d'entreprise intégrées permet de réduire le nombre des systèmes temporaires, qui augmentent les coûts cumulés et génèrent des données redondantes dans des silos isolés.

Un certain nombre d'organisations de pointe ont réussi grâce à cette approche pour trois raisons simples : l'ensemble du processus est ancré dans des objectifs stratégiques ; il s'agit d'une approche qui facilite le dialogue continu entre l'informatique et les utilisateurs ; il s'agit d'une approche qui améliore grandement les acquis organisationnels. Les avantages sont considérables, d'abord le dialogue permet généralement aux projets de se développer au-delà de l'idée originale d'exploiter le potentiel de la BI ; ensuite en progressant de réussite en réussite les entreprises créent un « modèle » qui s'affine en permanence et permet une amélioration constante ; enfin tout ceci stimule l'enthousiasme pour un développement continu de l'environnement BI en démontrant clairement que les besoins des utilisateurs sont effectivement satisfaits. Au final les entreprises peuvent concrétiser la vision d'un environnement BI pertinent fonctionnant de manière optimale.



## Approches analytiques de pointe, data mining ou fouille de données

Le data mining, la fouille de données ou toutes les approches analytiques de pointe représentent les formes les plus ambitieuses de l'Informatique Décisionnelle. Les techniques sous jacentes sont issues des Statistiques ou de l'Intelligence Artificielle, et certaines comme les Réseaux de Neurones, les Arbres de Décision, les Réseaux Bayésiens ou la Logique Inductive furent développées il y a plusieurs décennies dans des contextes très différents. Attention il ne s'agit pas de techniques informatiques, mais de processus de recherche automatique d'information dans un grand volume de données en mettant en œuvre toute technique appropriée.

Le postulat fondamental de ces approches est que dans toute base de données enregistrant les événements d'un processus économique, les données ne sont pas distribuées au hasard. Bien que ces données soient simplement constatées, le plus souvent collectées à des fins purement opérationnelles, on suppose qu'elles ont été générées par une réalité en grande partie déterministe, mais selon des processus que généralement on ne connaît pas. Les approches analytiques cherchent à mettre en évidence, décrire et permettre de reconstruire les effets de ces processus. Cette action d'identification des effets d'un processus à partir de données se caractérise par la construction d'un modèle, ensemble de règles, d'équations, de formules qui rendent compte de la distribution des données dans la base.

### Exemple de méthodologie de data mining

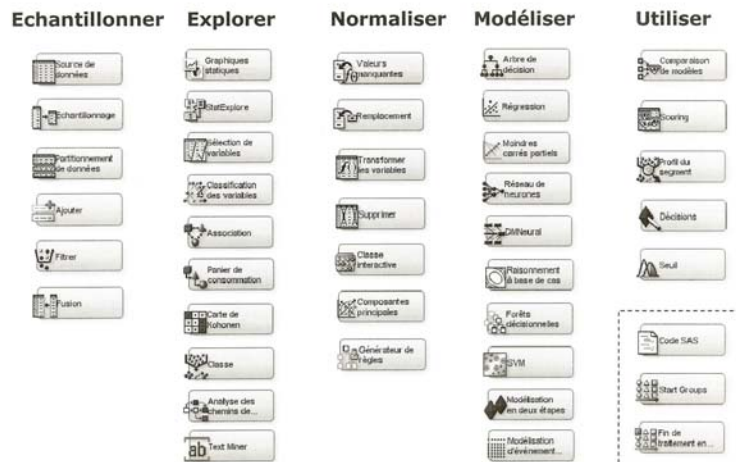


Illustration – 8

Construire méthodiquement un modèle des données dont on dispose sur un objet de gestion, est un saut important sur le chemin de la connaissance et de la décision. Au lieu de gigaoctets de données brutes, le décideur dispose alors d'une vue interprétable de son sujet d'étude. Mais attention à ne pas se laisser piéger par la métaphore liée à l'expression data mining qui suggère que les données (data) ne seraient qu'un amoncellement de débris, que l'on creuse (mining) à la recherche de la pépite. Elle fait croire que le data mining rejette une grande proportion des données disponibles, pour ne conserver qu'une petite partie

particulièrement significative. Or ce n'est pas exactement le cas, dans l'élaboration d'un modèle toutes les données sont prises en compte et traitées sensiblement sur un pied d'égalité. Le modèle est construit en fonction des régularités (patterns) de l'ensemble des données. Le succès de l'analyse se mesure à sa capacité à détecter les régularités fortement significatives.

D'autre part pour certains, toutes ces approches analytiques de pointe, data mining ou fouille de données ne seraient que de nouveaux noms branchés de la vénérable statistique. La réponse est : oui, mais .... Oui, car si vous faisons abstraction des techniques mises en œuvre, l'objectif est le même : élaboration et interprétation de modèles de la réalité construits à partir d'une description partielle de cette réalité par des données. Mais ..., car la statistique traditionnelle ne répond pas complètement aux attentes des utilisateurs potentiels. Le progrès le plus attendu des nouvelles approches c'est l'industrialisation des analyses, alors que la statistique reste une activité artisanale, réservée à un petit nombre de spécialistes traitant avec beaucoup de soin un petit volume de données très structurées, et mettant en œuvre un savoir faire peu automatisé. Or l'ambition des nouvelles approches c'est d'amener les biens faits de la statistique à tous les responsables sans qu'ils aient constamment recours à des statisticiens.

Cependant il ne faut pas se cacher, que la réalité est souvent compliqué, voire complexe, et que les approches analytiques peuvent être délicate à mettre en œuvre et leurs résultats parfois ambigus ou incertains, même si de nombreuses entreprises utilisent ces techniques avec beaucoup de succès dans toutes les fonctions de l'entreprise et en particulier dans les domaines de la relation client, de la logistique, de la qualité.

## **Cinq mythes sur le data mining**

De nombreuses entreprises florissantes ont découvert que les bruits qui circulent à propos du data mining ne se résument finalement qu'à de simples mythes. Plutôt que de céder à ces sirènes, ceux qui ont su voir plus loin y ont gagné un formidable avantage concurrentiel en utilisant le data mining pour résoudre des problèmes d'entreprise complexes et voir augmenter leur rentabilité.

Au fur et à mesure que les compagnies de télécommunications poursuivent ce petit « plus » qui permettra de relancer le revenu moyen par utilisateur (ARPU) et d'améliorer la rentabilité et l'efficacité de leurs opérations, le data mining et la transformation de données en informations directement utilisables tendent à se répandre toujours plus largement. Il reste néanmoins beaucoup à faire. Sous la pression accrue de la concurrence et des contraintes réglementaires, les compagnies de télécommunications cherchent à offrir un meilleur service pour un coût réduit. Fournir un meilleur service ne se résume pas à offrir une qualité de réseau optimale : cela implique aussi d'être en mesure de comprendre réellement les besoins des utilisateurs et d'y répondre à l'aide de services rentables. De même, réduire les coûts ne signifie pas simplement diminuer les prix. Dans la conjoncture économique actuelle, la pression des tarifs agit souvent comme un catalyseur déclenchant l'examen des systèmes, des processus et des coûts internes, ainsi que la recherche de l'excellence opérationnelle. Offrir un service meilleur et plus rentable, d'une part, et atteindre l'excellence opérationnelle, d'autre part, sont pour les entreprises deux objectifs pour lesquels la technologie du data mining peut constituer un atout précieux.

C'est d'ailleurs une technologie de data mining de pointe qui a convaincu les magasins Wal-Mart installés à la campagne de miser sur un type spécifique de Spam pour la saison de la chasse. Allez-y. Souriez. Il n'empêche que l'opération Spamouflage avec des boîtes en tenue de camouflage s'est révélée un énorme succès. Au-delà d'une idée simplement séduisante, Spamouflage a aidé Wal-Mart à tirer des revenus supplémentaires de ses clients existants et a démontré à quel point ce distributeur comprenait les gens à qui il s'adresse.

Bien qu'étranger au monde des télécommunications, cet exemple n'en reste pas moins significatif de la façon dont des entreprises leaders sur leur marché peuvent transformer des données en informations, puis en actions et, au bout du compte, en bénéfices.

Mais qu'est-ce donc que le data mining ?

Le data mining est un puissant outil d'analyse qui permet à des responsables d'entreprises d'aller plus loin que la simple description des comportements passés d'une clientèle et de prédire l'avenir. Il permet de déceler dans les comportements « mystérieux » des clients des règles qui les sous-tendent. Grâce à ces découvertes, il est possible d'augmenter les revenus, de réduire les dépenses, d'identifier des opportunités commerciales et donc de bénéficier d'avantages concurrentiels certains.

Le développement de mythes autour du data mining s'explique en partie par le fait que le concept reste encore flou pour bien des gens. À la base, le data mining se définit comme un jeu de techniques mathématiques complexes servant à découvrir et à interpréter des schémas jusqu'ici inconnus dans un ensemble de données détaillées. Depuis le milieu des années 1980, période à partir de laquelle le data mining a commencé à sortir du champ de la recherche universitaire, médicale et scientifique, ces techniques ont été appliquées avec une grande efficacité dans le domaine des télécommunications, de la distribution, des banques, des assurances, des transports et de l'hôtellerie.

La réputation d'outil analytique du data mining l'a souvent fait confondre avec la technique de traitement analytique en ligne nommée OLAP. L'OLAP est une technique analytique précieuse lorsqu'on l'applique à l'analyse d'opérations commerciales afin d'en tirer une perspective historique des événements. Imaginons par exemple qu'un directeur du marketing veuille comprendre pourquoi les ventes ont chuté dans une région donnée. Les outils d'OLAP lui permettent d'interroger la base selon de multiples critères, qu'il s'agisse des ventes par segment, par produit, par service ou de leur évolution. En examinant l'historique des données sous plusieurs angles, ce responsable sera en mesure d'identifier les causes (point de vente, produits ou période) qui ont pesé sur les ventes.

Le data mining s'intéresse à un ordre de problèmes différent. Il peut servir à prédire des événements futurs, comme les ventes du mois suivant en fonction des promotions ou le type de consommateur qui sera le plus sensible à une vente promotionnelle. La façon dont un certain nombre d'entreprises l'utilisent déjà permet de dissiper les cinq grands mythes qui planent sur le data mining.

Mythe numéro 1 : Le data mining pourrait instantanément prédire l'avenir, à la manière d'une boule de cristal

Le data mining n'est ni une boule de cristal ni une technologie capable de faire apparaître des réponses comme par magie en appuyant sur un bouton. C'est un processus qui comporte

plusieurs étapes : définition du problème de l'entreprise, exploration et préparation des données, développement d'un modèle et déploiement des connaissances obtenues. En général, les entreprises consacrent l'essentiel du temps nécessaire à pré-traiter et à préparer les données afin d'obtenir des Données nettoyées, cohérentes et convenablement combinées pour fournir les renseignements fiables qu'elles en attendent. La qualité des données est au cœur du data mining et un data mining réussi exige des données reflétant fidèlement l'activité de l'entreprise. Dans le cas des télécommunications, les opérateurs doivent comprendre que la puissance du data mining réside dans sa capacité à relever des défis spécifiques de nature prédictive ou descriptive. Il peut s'agir de : segmenter la clientèle, prédire la propension des clients à acheter (ou à changer de prestataire), détecter la fraude, augmenter l'efficacité de l'organisation (par exemple en optimisant l'assurance de revenus et les performances réseau).

Les entreprises de télécommunications qui comprennent comment fonctionne le processus en retirent des résultats indéniables. Tel opérateur européen de téléphonie mobile connaît par exemple la valeur de ses clients. Pas seulement les revenus qu'il peut en tirer, mais leur valeur elle-même.

Les revenus et les coûts sont calculés pour chaque client, les programmes de marketing et de GRC (gestion de la relation client) sont ensuite élaborés en fonction de la rentabilité de ces clients. Telle compagnie de télécommunication sud-américaine anticipe et intervient pour empêcher le départ de clients à valeur élevée en repérant les schémas qui conduisent à la perte de clientèle, s'appuyant pour cela sur l'analyse des mesures d'utilisation, d'achat et de qualité de services. Tel autre opérateur européen de téléphonie mobile a utilisé le data mining pour analyser le risque de churn (tendance du consommateur à changer de prestataire), afin de prendre des mesures proactives ciblées visant à identifier les clients et les segments ayant la plus forte propension au churn. Armé de cette information, l'opérateur a lancé des campagnes de marketing ciblées qui lui ont permis de réduire de 50 % le taux de churn sur les segments concernés. En parallèle, la compagnie a bénéficié d'un taux d'augmentation de 30 % au cours de la campagne marketing, améliorant du même coup la satisfaction des clients et par conséquent leur durée de fidélité. Mais le data mining peut aller bien plus loin qu'une simple analyse du comportement de la clientèle. Un des principaux opérateurs aux USA utilise le data mining pour améliorer les performances de son réseau, sans avoir à dépenser des millions de dollars pour actualiser son infrastructure réseau existante. Grâce à une surveillance constante des règles de performances et à une analyse permanente de l'historique de l'utilisation des composants et des lignes principales, complétées par des mesures régulières de l'activité de son réseau, cet opérateur peut veiller à ce que les appels soient acheminés en utilisant la capacité disponible. Ces efforts lui permettent d'enregistrer une baisse des défauts de service et des interruptions techniques et, parallèlement, une hausse du nombre d'appels réussis passés sur son réseau. Non seulement le degré de satisfaction des utilisateurs augmente, mais l'opérateur est aussi en mesure de fournir la qualité et la disponibilité de service exigées (et mesurées) par les instances de régulation.

Mythe n° 2 : le data mining ne serait pas encore viable pour des applications professionnelles

Le data mining est une technologie viable et ses résultats professionnels sont hautement prisés. Ceux qui propagent le mythe en question sont en général les mêmes qui se trouvent contraints d'expliquer pourquoi ils n'utilisent toujours pas cet outil. Leur discours s'articule autour de deux arguments liés. D'après le premier : « Il est impossible d'utiliser efficacement de grandes bases de données ». Quant au second, il stipule que : « Le data mining ne peut



pas s'effectuer dans le moteur de l'entrepôt de données ». Ces deux points ont certes été vrais à une certaine époque. Mais n'oublions pas qu'il a longtemps été vrai aussi que les avions ne pouvaient pas quitter le sol.

Répondons simultanément à ces deux arguments. Les bases de données sont aujourd'hui tellement vastes, que les entreprises craignent que l'architecture informatique supplémentaire nécessaire aux projets de data mining n'entraîne d'énormes coûts, et que le traitement des données exigé pour chaque projet ne soit trop long. C'est oublier que certaines des bases de données modernes utilisent désormais une technologie de traitement en parallèle, qui autorise l'extraction de données au sein de la base. L'extraction de données dans la base permet aux entreprises de supprimer les déplacements de données, d'exploiter les performances du traitement en parallèle, de réduire la redondance des données et d'annuler les coûts de création et de maintenance d'une base de données distincte et redondante dédiée au data mining. La combinaison de l'extraction des données dans la base de données et du traitement en parallèle donne une technologie de data mining parfaitement viable.

La vraie question ne concerne donc pas la viabilité de la technologie du data mining mais consisterait plutôt à se demander : " Comment le data mining peut-il m'aider à améliorer l'activité de mon entreprise ? " Dans le secteur des télécoms, le data mining est une pratique désormais éprouvée qui a donné de vrais résultats avec un retour sur investissement rapide dans de nombreux secteurs, notamment la relation client, la gestion des campagnes de marketing, la mesure de la valeur client, la gestion financière ou l'optimisation de l'assurance de revenus et des performances réseau, pour n'en citer que quelques-uns.

Mythe n° 3 : le data mining exigerait une base de données distincte et dédiée

A en croire les prestataires de data mining, il serait indispensable de disposer d'outils onéreux et dédiés (base de données, data mart ou serveur analytique) pour mettre en place cette technique. Ils invoquent pour cela la nécessité d'extraire les données sous un format propriétaire pour un traitement efficace. Or, non seulement ces data marts sont coûteux à l'achat et à la maintenance, mais ils exigent que les données soient extraites pour chaque projet distinct de data mining, ce qui représente une procédure coûteuse et fortement consommatrice de temps.

Heureusement, les progrès réalisés dans le domaine des bases de données n'exigent plus d'effectuer le data mining dans un data mart distinct. En réalité, pour une opération de data mining efficace, il faut pouvoir disposer d'un entrepôt de données à l'échelle de l'entreprise, ce qui, au regard du coût d'investissement total, se révèle considérablement moins cher que d'utiliser des data mart distincts.

Voici pourquoi. Au fur et à mesure que les sociétés mettent en place des projets de data mining dans l'ensemble de l'entreprise, le nombre d'utilisateurs exploitant les modèles de data mining ne cesse de croître, de même que les besoins d'accès à de larges infrastructures de données. Un entrepôt de données d'entreprise d'avant-garde ne se contente pas de stocker efficacement l'ensemble des données de l'entreprise et de rendre largement inutiles d'autres data marts ou d'autres entrepôts. Il constitue également les fondations idéales pour des projets de data mining, en fournissant un répertoire de données unique à l'échelle de l'entreprise, capable d'offrir une vision cohérente et actualisée de la clientèle. En outre, l'intégration d'extensions de data mining au sein de l'entrepôt de données permet à l'entreprise de réaliser deux types d'économies supplémentaires. Tout d'abord, il n'est plus

nécessaire d'acheter et d'assurer la maintenance de matériel supplémentaire uniquement dédié au data mining. Ensuite, les entreprises voient se réduire la nécessité d'effectuer un va-et-vient de données depuis et vers l'entrepôt pour réaliser les projets de data mining, ce qui, nous l'avons vu, est une opération gourmande en travail et en ressources.

Par exemple, l'un des plus grands opérateurs de téléphonie mobile aux USA utilise un EDW (Enterprise Data Warehouse, entrepôt de données d'entreprise) centralisé pour fournir des informations à une large gamme d'applications commerciales, depuis l'attention à la clientèle jusqu'au marketing. Il n'en a pourtant pas toujours été ainsi. Cet opérateur, desservant de nombreux marchés régionaux, a été conduit à établir différents data marts régionaux. Obtenir une vue holistique des informations à l'échelle de l'entreprise était devenu pour le moins difficile. Les prises de décision souffraient de l'absence de données ou d'informations cohérentes, ce qui affectait le moral d'une équipe qui ne pouvait pas visualiser le fruit de son travail.

En transférant le tout vers un EDW centralisé, cet opérateur a pu constater des résultats immédiats. Les données étaient devenues cohérentes, les décisions pouvaient être prises en toute sécurité, avec l'assurance que les données reposaient sur des fondations saines. En termes de performances, le nouvel EDW délivrait des informations cohérentes presque 90 % plus vite qu'avec l'ancienne approche par data marts fragmentés. Lorsqu'il est nécessaire d'analyser de grandes quantités de données, de tels gains de performances font la différence entre être présent le premier sur un marché en se fiant à son seul instinct, et s'y présenter en sachant que les décisions prises l'ont été sur la base d'informations tirées du comportement réel des clients.

Mythe n° 4 : il faudrait être polytechnicien pour faire du data mining

Certains estiment que le data mining est une technologie si complexe qu'il faut au moins trois polytechniciens pour la maîtriser : un spécialiste en statistiques ou en méthodes quantitatives, un commercial fin connaisseur de la clientèle et un informaticien.

La vérité oblige à dire que bien des projets réussis ont été menés à bien sans l'aide d'un seul de ces polytechniciens. On peut citer l'exemple d'une société sud-américaine de télécommunications qui a réussi à repérer des changements dans le comportement de ses clients, ce qui lui a permis de conserver 98 % de ses clients à valeur élevée en pleine période de dérégulation. Ce succès est à mettre au crédit d'une équipe multidisciplinaire travaillant en collaboration.

Le data mining demande un effort de collaboration de la part d'un personnel qualifié dans trois domaines. Les commerciaux doivent guider le projet en créant une série de questions commerciales spécifiques, puis interpréter les résultats obtenus. Les concepteurs de modèles analytiques, formés aux techniques, aux statistiques et aux outils du data mining, doivent construire un modèle fiable. Les informaticiens doivent apporter leurs compétences dans le traitement et la compréhension des données, ainsi que leur indispensable soutien technique.

Mythe n° 5 : le data mining serait réservé aux grandes entreprises disposant d'un large volume de données client

La réalité est que toute entreprise, quelle que soit sa taille, possédant des données qui reflètent avec précision son activité où sa clientèle peut s'en servir pour construire des modèles et en tirer des informations précieuses sur ses défis commerciaux majeurs. Le volume des données clients d'une société n'a jamais été en cause.

Par exemple, un opérateur de taille moyenne de téléphonie mobile dans la région Asie/Pacifique se trouvait confronté à un problème de fraude. Il a fait appel à une base de données centralisée pour analyser plus vite l'historique détaillé de ses appels et pour mieux comprendre le comportement de sa base client. La détection des utilisations et des abonnements frauduleux n'a ainsi pris que quelques jours au lieu de plusieurs mois et l'opérateur a pu constater une diminution des pertes dues à la fraude de 60 %. À l'heure où les opérateurs subissent une pression croissante les poussant à augmenter leur nombre d'abonnés et à améliorer le revenu moyen par utilisateur, la possibilité d'agir positivement sur le bilan sans avoir à conquérir de nouveaux abonnés ni à augmenter l'ARPU a eu des conséquences bénéfiques notables sur les performances financières de l'opérateur en question. Bien qu'il ait dû créer la base de données pour traiter spécifiquement son problème de fraude, l'opérateur en a rapidement tiré bien d'autres profits. Les économies réalisées grâce aux opérations informatiques et à l'intégration des systèmes ont permis de raccourcir les cycles de développement d'applications en exploitant les données provenant de l'entrepôt de données, et de réduire les coûts de 30 à 50 %. Une fois ces bénéfices constatés, des départements de plus en plus nombreux se sont mis à utiliser la base de données, permettant ainsi l'accès rapide aux informations commerciales stratégiques et de promptes prises de décisions dans des domaines tels que le churn, les activations quotidiennes et les flux de revenus. Les utilisateurs ont pu constater que des informations qui n'étaient auparavant accessibles qu'au bout de plusieurs semaines ou plusieurs mois étaient désormais accessibles de façon quotidienne.

Profitez-en

En conclusion : le data mining n'est plus aujourd'hui ni lent, ni coûteux ni d'utilisation trop complexe. La technologie et le savoir-faire commercial existent et permettent de mettre en place un processus efficace et économe. Des compagnies de télécommunications de tailles diverses font partie des entreprises qui ont décidé de mettre les vieux mythes à l'épreuve et de prouver que le data mining est un processus essentiel pour pouvoir prospérer dans un univers professionnel dominé par une concurrence féroce et par l'obsession du client.

### **En analyse de données comme en cuisine, la réussite se joue en grande partie avec les ingrédients**

Tout cuisinier sait que la réussite d'un chef-d'œuvre culinaire, comme des macarons, réside en grande partie dans les ingrédients. Il est difficile de préparer un excellent repas si les ingrédients sont mauvais ou si leur combinaison n'est pas adéquate. Il en va de même pour l'analyse de données. Si les données sont incomplètes, inexactes ou sans rapport avec le problème à résoudre, il sera difficile, voir impossible, de créer un modèle. Par exemple, si le modèle de calcul de la valeur des clients attribue un faible score à certains clients rentables car les transactions en ligne ou les commandes spéciales ne sont pas prises en compte, il y a le risque de perdre certains des meilleurs clients. L'efficacité d'un modèle d'analyse de données est donc directement proportionnelle à la qualité des données. En d'autres termes, on ne peut pas faire de bons macarons avec de mauvais ingrédients.

L'analyse de données exploite des algorithmes de statistiques et d'apprentissage automatique pour trouver des informations qui peuvent contribuer à résoudre les problèmes quotidiens que rencontrent les entreprises. Lorsque les utilisateurs font de l'analyse de données, ils mettent en œuvre en général des algorithmes mathématiques tels que des réseaux neuronaux, des arbres de décision et d'autres techniques statistiques complexes qui servent à rechercher des tendances dans les données. S'il est vrai que ces algorithmes sont une partie importante de l'analyse de données, il convient de noter que ces outils recherchent des tendances dans n'importe quelles données, quelle que soit les capacités de ces dernières à représenter les comportements et les tendances que l'on essaye de modéliser. Pour cette raison, la préparation des données est l'une des étapes les plus critiques dans l'analyse de données et pourtant, il s'agit souvent de l'une des étapes les plus négligées.

La première étape dans la préparation des données consiste à recueillir les données relatives au problème à résoudre. Si un utilisateur possède un entrepôt de données d'entreprise, le processus est considérablement simplifié. Au contraire, si les données sont stockées en divers endroits, il faut explorer plusieurs sources afin d'identifier les données disponibles pour résoudre le problème. Dès que les données qui doivent être analysées sont définies, il convient de les intégrer, les évaluer et éventuellement les transformer pour s'assurer qu'elles sont valides d'un point de vue conceptuel, cohérentes et statistiquement analysables. Par exemple, si les données proviennent de différentes sources, il faudra résoudre de nombreux problèmes de formats et de définitions.

Même si un utilisateur a la chance d'avoir un entrepôt de données d'entreprise, il est possible que les données qu'il contient ne conviennent pas en l'état pour l'analyse envisagée. Il est indispensable alors d'isoler et de préparer les données pour le modèle. Cela signifie qu'il faut travailler en collaboration avec les analystes et les experts en données afin de définir les éléments qui sont nécessaires pour réaliser le modèle. Il est primordial, pour chaque variable, de définir s'il faut utiliser toutes les données ou uniquement un sous-ensemble. Il convient aussi de définir une stratégie pour traiter les valeurs aberrantes (données hors norme) ou éventuellement développer quand même un modèle reposant sur ces valeurs. Par exemple, si l'objectif est de prévoir les taux de fréquentation et les revenus de rencontres sportives, il faut certainement éliminer les chiffres de fréquentations anormales dues à des événements particuliers, grève des transports, etc.... Au contraire dans le cas de la détection de fraudes, il peut-être pertinent de se concentrer sur certaines valeurs aberrantes car elles sont peut-être la représentation de transactions frauduleuses.

Une fois que les données sont sélectionnées, il convient de les analyser à l'aide de techniques de statistiques descriptives et de visualisation pour identifier les problèmes de qualité et mieux comprendre les caractéristiques des données. Des problèmes de qualité des données peuvent être mis en lumière, tels que des valeurs manquantes qui peuvent nuire à l'intégrité de n'importe quel modèle d'analyse. Il faut alors compenser et corriger les problèmes identifiés. Ainsi, s'il manque des données, il faut déterminer la meilleure méthode pour abandonner ou remplacer ces valeurs manquantes. Certaines techniques d'analyse de données permettent d'estimer les valeurs manquantes sur la base d'autres valeurs mesurées.

Il existe de nombreuses techniques qui peuvent être employées pour obtenir de meilleurs modèles. Il s'agit par exemple de créer des variables "dérivées", de remplacer des valeurs manquantes ou d'utiliser des techniques d'agrégation ou de réduction des données. Il peut être nécessaire de rechercher les meilleurs agrégats ou de nouvelles variables analytiques

pour construire un modèle optimum. Par exemple, dans le cadre de la préparation des données relatives aux clients en vue d'un programme de marketing pour un nouveau prêt, le rapport endettement/revenu peut être un indicateur mieux adapté que le seul revenu ou le seul endettement.

Enfin, les données doivent être transformées dans un format adapté aux algorithmes d'analyse. De nombreux algorithmes d'analyse de données requièrent la transformation des données de classification (non numériques) en données numériques ou la réduction de celles-ci dans une plage particulière. Certains algorithmes et techniques statistiques nécessitent également que les données numériques possèdent des propriétés spécifiques qui n'existent peut-être pas dans les données avant la transformation. Pour ces variables, il faudra peut-être les encoder à nouveau ou les transformer pour produire les variables adéquates pour les techniques d'analyse de données. Ainsi la valeur des données est directement proportionnelle au temps et au soin consacré à leur préparation en vue de régler un problème analytique particulier. A l'instar de ce que dirait un pâtissier, la qualité du résultat final dépend en grande partie des ingrédients.

Comme il l'a été dit plus haut, la préparation des données est un processus qui peut prendre du temps. En fait, les spécialistes de l'analyse de données qui comprennent l'importance de la préparation peuvent passer jusqu'à 70 pour cent de l'ensemble du processus d'analyse, à l'étude et au prétraitement des données. Ils agissent ainsi du fait du caractère nécessairement itératif de la préparation des données. Lors de l'étude et du traitement préalable, l'analyste réalise des expériences avec les données. Il examine les relations entre les variables et ceci peut nécessiter plusieurs itérations de requêtes ad hoc ou d'autres analyses sur un volume important de données. Jusqu'il y a peu cette préparation était difficile car la majorité des outils d'analyse de données ne fournissait pas les fonctions nécessaires pour effectuer avec efficacité l'étude et le traitement de large volume de données, et ceci était un inconvénient majeur pour les spécialistes de l'analyse de données.

Les personnes qui savent que ces tâches de préparation des données sont nécessaires et très utiles, mettent en œuvre des moyens techniques compliqués, comme l'analyse de grands volumes de données. La majorité d'entre elles prépare les données directement dans les bases afin d'éviter de devoir déplacer de grands volumes de données vers un serveur ou un poste de travail dédié à l'analyse de données. Cette méthode a fait ses preuves et les spécialistes de l'analyse de données ont souvent utilisés des programmes SQL qu'ils ont fait spécifiquement développés ou bien ils ont appris eux-mêmes le langage SQL. Ceci étant dit, l'on observe une évolution nette en la matière vu les progrès réalisés dans les technologies de gestion de données.

Il existe désormais des alternatives aux programmes SQL développés « manuellement » pour la préparation des données. De nouvelles solutions telles que les technologies de génération SQL au sein des bases de données permettent de réaliser la préparation des données directement au sein de la base d'une entreprise. Au lieu de déplacer les données vers l'outil, des solutions du type Analytical Data Set Generator (Générateur de jeux de données analytiques) permettent de placer les fonctions nécessaires et les traitements directement au sein d'une base de données. Ces solutions accélèrent non seulement l'analyse grâce à l'utilisation du moteur de base de données qui prend en charge toutes les fonctions d'analyse, mais ils offrent également des fonctions qui facilitent et dans certains cas automatisent des tâches de prétraitement des données. Les fonctions d'exploration de données sont

particulièrement efficaces au niveau du profilage des données grâce à une intelligence intégrée qui analyse et transforme les données.

Les modules de création de variables et de transformation de variables au sein d'un ADS Generator favorisent le recours aux meilleures pratiques en matière de préparation des données. Les interfaces conviviales permettent d'analyser rapidement et efficacement les données de plusieurs tables. Les résultats de ces analyses sont présentés dans des tableaux et des graphiques faciles à interpréter. Cela signifie que l'étude et le prétraitement des données, qui nécessitaient des semaines de travail, peuvent être désormais réalisés en quelques jours, voire en quelques heures. Cela reviendrait à pouvoir préparer de meilleurs macarons en un temps record... peut-être même en quelques minutes. Le plus grand des pâtisseries serait comblé.

Les technologies de préparation des données telles que l'ADS Generator font que l'étude et le prétraitement des données ne prennent plus autant de temps que par le passé. Au contraire, grâce à ces nouvelles technologies, la préparation des données est devenue beaucoup plus simple, plus facile et plus rentable. Pensez-y la prochaine fois que vous mangerez un macaron.

### **Fouille de Données : une démarche en cinq étapes !**

Pour mettre en place une approche de fouille de données, il ne s'agit pas uniquement de choisir la bonne technique statistique. Il faut bien plus qu'une bonne technique statistique. Les spécialistes ont développé un cadre formel de fouille de données qui couvre en particulier deux activités primordiales de ce type d'approche : la gestion de projet et le transfert des connaissances. La fouille de données doit être organisée selon cinq étapes consécutives : identification des problèmes de l'activité, préparation de l'architecture, préparation des données, analyse et livraison des connaissances.

Durant la première étape, il convient d'identifier, de préciser et de qualifier les questions pour la fouille de données. Il s'agit d'identifier aussi la meilleure démarche à mettre en œuvre pour répondre aux questions sachant qu'il existe de nombreuses techniques d'analyse commerciale, et que parmi celles-ci la fouille de données est celle qui nécessite le moins de ressources. Il faut se poser les questions suivantes pour définir si la fouille de données convient : Les réponses aux questions nécessitent-elles la compréhension de relations complexes entre plusieurs variables ? Y a-t-il des questions pré-requises auxquelles il faut répondre ? Quelles sont-elles ? Les réponses aux questions nécessitent-elles des déductions relatives à un phénomène ou à l'avenir ? Les réponses décriront-elles des phénomènes complexes ? Quel est le niveau de technologie requis pour gérer la complexité des réponses aux questions ?

Durant la préparation de l'architecture, il convient d'étudier les aspects techniques de la démarche et d'élaborer un plan. Par exemple, il faut envisager et planifier les implications techniques de l'ajout d'un modèle d'analyse et de nouvelles infrastructures à l'architecture logicielle et matérielle actuelle. Il faut aussi identifier les logiciels d'analyse qui conviennent pour répondre aux questions définies, fixer les rapports entre les modèles d'analyse et d'autres applications, identifier le besoin d'autres technologies, mettre en lumière les contraintes et les limites au niveau de l'infrastructure, décrire les caractéristiques de l'environnement de production requis pour utiliser et maintenir les modèles analytiques, et

enfin définir tous les problèmes de transport de données entre le modèle analytique et les environnements de production.

Lors de la préparation des données, il faut en outre identifier, extraire et valider de grands échantillons de données, les déplacer dans l'environnement d'analyse, tester leur pertinences par rapport aux problèmes à résoudre et enfin développer, affiner les modèles préliminaires afin de garantir des résultats encore plus solides.

La préparation des données n'est pas une tâche insignifiante. Elle représente en général 70% du délai. Les entreprises qui ont déjà nettoyé et transformé leurs données dans un entrepôt ont une longueur d'avance, mais il leur reste quand même du travail. En effet, les seuils de qualité des données pour la fouille de données sont bien plus élevés que ceux pour les utilisations habituelles d'un entrepôt de données. Par exemple, il est nécessaire de passer les données en revue à la recherche des valeurs nulles qui sont inacceptables dans le cadre de certaines techniques d'analyse. Une fois que les valeurs nulles sont connues, il faut décider soit de remplacer ces valeurs nulles par des valeurs raisonnables et utiles ou soit tout simplement de supprimer les enregistrements correspondants. Il faut évaluer également le caractère variable de chaque élément pour s'assurer que les données ont suffisamment changé pour garantir des inférences fiables et valides. Enfin, alors que les premières informations sélectionnées sont mieux connues il faut explorer d'autres données qui pourraient venir les compléter. Il n'est pas rare de tester de grands échantillons de données avant de trouver la sélection de données pertinentes pour répondre aux objectifs et aux critères de qualité.

Ensuite il convient de passer à la phase d'analyse qui recouvre le développement, l'essai et la validation des modèles analytiques. La modélisation analytique marque le début des choses sérieuses et à la fin de cette étape les réponses aux questions que se posait l'entreprise sont données et documentées. Lorsqu'il a été vérifié que le modèle fournit des résultats valides et utiles, l'application pour les utilisateurs peut être développée afin que les modèles puissent être intégrés dans les processus habituels de l'entreprise. A ce stade il convient d'organiser plusieurs séances de formation pendant la livraison des connaissances destinées à ceux qui utiliseront, entretiendront et rafraîchiront les modèles.

La durée moyenne d'un premier projet de fouille de données est de trois mois (4 semaines pour certains et 6 mois pour d'autres). La complexité des problèmes, les exigences au niveau de l'architecture et des technologies, l'importance de la préparation des données, la complexité des analyses et l'ampleur du transfert de connaissances influencent la durée d'un projet de fouille de données.

Tant tout les cas il est impératif de très bien définir le problème à résoudre et de s'assurer que la fouille de données est la solution la plus adaptée. Les questions les mieux traitées grâce à la fouille de données sont celles qui cherchent à expliquer un phénomène relativement complexe ou qui ont une inférence sur des événements ou un comportement futur. Les questions qui nécessitent une synthèse ou une simple description à l'aide de quelques variables sont traitées plus efficacement avec les techniques traditionnelles de prises de décision et les statistiques descriptives.

Pour les projets de fouille de données, il convient de sélectionner des individus qui sont curieux, qui font preuve d'une capacité de réflexion analytique et qui ont d'excellentes aptitudes en technologies de l'information. Un diplôme en statistiques ou en mathématiques

appliquées est souhaitable, mais pas toujours indispensable. D'autre part, une participation active des experts de l'activité et des utilisateurs des résultats de la fouille de données est critique pour la réussite de ce type de projet. Les individus qui ont travaillé en tant qu'analystes d'entrepôt de données ou qui ont assuré la liaison entre les spécialistes de l'activité et les spécialistes des technologies de l'information possèdent de nombreuses aptitudes au niveau de la résolution de problèmes, et jouissent de l'expérience nécessaire pour assurer la réussite des projets de fouille de données. Les collaborateurs expérimentés qui ont développé des entrepôts de données et qui connaissent leur utilisation complexe sont d'excellents candidats, car ils connaissent déjà le sens que revêt l'entrepôt de données pour l'entreprise et le pourquoi et le comment des éléments développés. L'apprentissage des techniques analytiques et statistiques requises pour un projet de fouille de données particulier prend bien moins de temps que la compréhension en profondeur du contenu et de la logique de la base de données.

Enfin d'un point de vue technique, il est préférable de réaliser les qualifications, les examens et les analyses des données au sein de l'entrepôt de données autant que possible. Comme cela a été dit plus haut, l'identification de données pertinentes et d'excellente qualité occupe la plus grande partie du temps. Cela s'explique pour deux raisons : tout d'abord, lorsque des données ne satisfont pas aux critères de qualité, il faut revenir dans la base de données afin d'y rechercher d'autres éléments à évaluer. Plus vous en apprenez sur les données, plus vous découvrez de nouvelles options et de nouveaux points de vue qui peuvent requérir un raffinement de l'approche analytique et la sélection d'autres données. Globalement ce processus implique de nombreuses itérations et tout cela est facilité si les opérations se font au sein de l'entrepôt de données..

### **Fouille de Données : les biens faits des approches en laboratoire !**

Tous ceux qui envisagent la fouille de données à un certain moment partagent une même caractéristique, ils s'interrogent sur de nombreux points et notamment sur l'intérêt d'une telle approche. Vu que généralement nous apprenons grâce aux exemples, la plupart de ces interrogations portent sur ce que font les autres spécialistes de la fouille de données : quels problèmes résolvent-ils à l'aide des technologies de fouille de données ? Comment s'y prennent-ils ? Combien de temps leur faut-il ? En retirent-ils quelque chose ? Dans le but d'aider les entreprises qui étudient la valeur de la fouille de données, des sociétés de conseil ont mis sur pied des laboratoires de fouille de données. Ils offrent aux entreprises le cadre, les conseils et les outils pour tester la fouille de données et réaliser des projets pilote avant de mettre en œuvre leurs propres solutions.

Par exemple, il s'agit de se focaliser sur une ou deux questions spécifiques relatives au comportement des clients et en particulier pouvoir les catégoriser. Ainsi, une grande banque internationale voulait contrôler les coûts générés par ses clients lorsqu'ils utilisaient les distributeurs automatiques d'autres banques. La société voulait obtenir une réponse pour chacune des quatre questions suivantes : Qu'est-ce qui constitue une utilisation excessive par le client des distributeurs automatiques de la concurrence ? Quels sont les clients qui génèrent des coûts excessifs par l'utilisation des distributeurs automatiques de la concurrence ? Quelle est la valeur qu'ils représentent pour notre banque ? A quoi devons-nous prêter attention lorsque nous utilisons ces résultats ?



La première découverte fut que 10% des clients de la banque généraient 90% des coûts des distributeurs automatiques. Cette constatation aurait peut-être pu être faite à l'aide de moyens traditionnels. Cependant, grâce à la fouille de données, il a pu être mis en évidence le fait que sur les 10% des clients qui généraient les coûts, 80% étaient des clients de faible valeur. Ensuite à l'aide de plusieurs techniques de fouille de données permettant de prendre en compte de multiples variables, il a été possible de comprendre la valeur potentielle de chacun des clients de faible valeur et quantifier le concept d' « utilisation excessive ». La fouille de données a également permis de répondre à la question logique suivante que devait se poser la banque : fallait-il revoir le service offert à ces 80% de clients de faible valeur ? L'analyse a montré qu'environ 30% de ces clients de faible valeur étaient des clients à fort potentiel, à savoir : des étudiants. Ce n'était certes pas une découverte surprenante mais elle aurait été difficilement réalisable sans l'analyse de multiples variables.

La découverte la plus intéressante réalisée dans le cadre de la résolution du problème du coût des distributeurs automatiques fut qu'un concurrent ciblait les campus universitaires en vue d'y installer de nouveaux distributeurs automatiques. Aucune autre banque ne développait d'actions sur les campus et celle-ci jouissait d'une présence quasiment exclusive. Grâce à cette expérience en fouille de données, la banque pu répondre à ses questions initiales. Mais ce qui est encore plus important, c'est qu'elle fut en mesure de découvrir la stratégie d'un concurrent.

L'intérêt suscité par la fouille de données varie en fonction des secteurs. Il semblerait que le marketing et les finances marquent le plus d'intérêt pour la fouille de données. Les spécialistes du marketing doivent maîtriser le retour sur investissement de leurs opérations et pour cela, ils cherchent à cibler les campagnes sur les clients les plus susceptibles d'acheter un produit ou un service particulier. Les départements financiers sont intéressés par la possibilité d'aligner le coût des services sur les revenus (ou les revenus potentiels) tirés d'un client ou d'un segment de clients. Parmi les principales priorités de ces groupes, citons la compréhension et la réduction de l'attrition.

Exemple d'un cas classique de modélisation des réponses marketing. Une entreprise de télécommunication obtenait un taux de retour de 0,5% sur ses campagnes marketing. Elle souhaitait découvrir les clients les plus susceptibles d'acheter des services groupés. Cette entreprise envisageait le développement de capacités en matière de fouille de données mais elle voulait d'abord voir comment la fouille de données fonctionnait. Avec l'aide d'une société spécialisée, un modèle de propension à l'achat a été créé afin d'aider l'entreprise à prévoir les services qui, une fois groupés, seraient susceptibles d'être achetés par des clients particuliers. Le déploiement du modèle dans un territoire de vente test s'est traduit par une augmentation d'un facteur 10 des ventes de services groupés. Ces résultats démontrèrent l'exactitude des prévisions du modèle et convainquirent l'entreprise de l'efficacité de la fouille de données. Les résultats de ce test restreint furent tels que l'entreprise a depuis lors appliqué le modèle au niveau national. L'entreprise a également suffisamment appris sur les procédures de fouille de données que la société spécialisée utilise, et a suffisamment développé ses aptitudes au niveau des outils d'analyse qu'elle développe aujourd'hui ses modèles de manière indépendante.

## Faire son Data Mining directement dans son Entrepôt de Données

Le Data Mining est une puissante technique d'analyse qui par exemple permet à des responsables d'entreprises d'aller plus loin que la simple description des comportements passés d'une clientèle et de prédire l'avenir. Il permet de déceler dans les comportements « mystérieux » des clients des règles qui les sous-tendent. Grâce à ces découvertes, il est possible d'augmenter les revenus, de réduire les dépenses, d'identifier des opportunités commerciales et donc de bénéficier d'avantages concurrentiels certains.

Mais à en croire certains fournisseurs de solutions et de prestations de data mining, il serait indispensable de disposer d'outils dédiés (base de données, data mart ou serveur analytique) pour faire du data mining. Ils invoquent pour cela la nécessité d'extraire les données sous un format propriétaire pour réaliser des traitements efficaces. Or, non seulement ces moyens spécialisés sont onéreux à mettre en place et à maintenir, mais ils exigent que les données soient extraites pour chaque projet distinct de data mining, ce qui représente une procédure coûteuse et fortement consommatrice de temps. Heureusement, les progrès réalisés dans le domaine des bases de données n'exigent plus d'effectuer le data mining dans des data marts distincts. En réalité, pour une opération de data mining efficace, il suffit de s'appuyer sur son entrepôt de données, ce qui, au regard du coût d'investissement total, se révèle considérablement moins cher que d'utiliser des moyens spécifiques.

### Construction et déploiement de modèles dans l'entrepôt de données

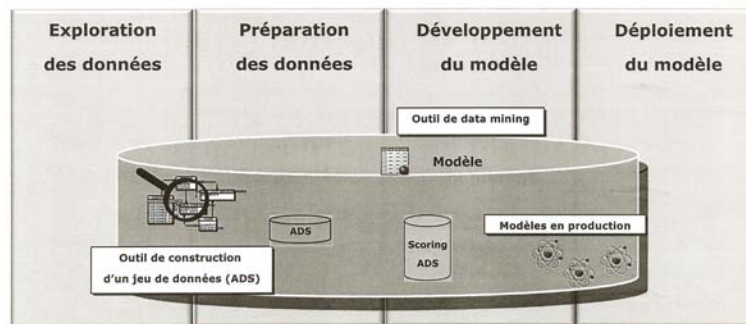


Illustration – 9

Au fur et à mesure que les sociétés mettent en place des EDW (Enterprise Data Warehouse, entrepôt de données d'entreprise) et couvrent l'ensemble des besoins décisionnels des différentes divisions et fonctions de l'entreprise, le nombre d'utilisateurs exploitant des modèles de data mining ne cesse de croître. En effet un EDW bien architecturé ne se contente pas de stocker efficacement l'ensemble des données historiques, il rend inutiles d'autres data marts ou d'autres moyens spécialisés de stockage. Un EDW constitue la fondation idéale pour des projets de data mining, en fournissant un répertoire de données unique à l'échelle de l'entreprise, capable d'offrir une vision cohérente et actualisée des activités. En outre, l'intégration de fonctionnalités de data mining au sein de l'EDW permet à

l'entreprise de réaliser deux types d'économies supplémentaires. Tout d'abord, il n'est plus nécessaire d'acheter et d'assurer la maintenance d'un matériel complémentaire uniquement dédié au data mining ; ensuite les entreprises ne sont plus obligées d'effectuer un va-et-vient de données depuis et vers l'entrepôt pour réaliser les projets de data mining, ce qui, nous l'avons vu plus haut, est une opération gourmande en temps et en ressources.

A titre d'exemple, l'un des plus grands opérateurs de téléphonie mobile aux USA utilise aujourd'hui un EDW pour fournir des informations à une large gamme d'applications commerciales et marketing. Cet opérateur, desservant de nombreux marchés régionaux, avait été conduit à établir différents data marts régionaux, et obtenir une vue d'ensemble des informations à l'échelle de l'entreprise était devenu particulièrement difficile. Les prises de décision souffraient de l'absence de données ou d'informations cohérentes, ce qui affectait les performances et le moral des équipes qui ne pouvaient pas anticiper et mesurer le fruit de leur travail. En transférant toutes ses données vers un EDW, cet opérateur a pu constater des résultats immédiats. Les données étaient mises en cohérence, les décisions pouvaient être prises en toute sécurité, avec l'assurance que les données reposaient sur des fondations saines. En termes d'efficacité, le nouvel EDW délivre des informations d'une meilleure qualité 90% plus vite qu'avec l'ancienne approche par data marts. Lorsqu'il est nécessaire d'analyser de grandes quantités de données, de tels gains permettent de faire la différence sur le marché.

### **Fouille de Données et PMML : vers une extraction de données plus rapide, plus facile et moins coûteuse**

Si une société perd 25 millions de dollars par mois, l'extraction des données peut prendre un caractère très urgent. C'est ce qu'endurent de nombreuses grosses sociétés de télécommunications sans fil – elles subissent un taux de rotation moyen du secteur de 2,4%, ce qui se traduit par une perte mensuelle de plusieurs centaines de clients. Avec un revenu mensuel moyen de 45€ par client, ces défections coûtent plus de 250 millions d'€ par an aux grosses sociétés. L'extraction des données permet aux fournisseurs de télécommunications sans fil d'atténuer ces pertes. En appliquant des méthodes d'analyse avancées à ce problème, un fournisseur peut collecter des informations, des prévisions, des descriptions, des notations et des profils. Grâce à tout cela, le prestataire de services peut prévoir quels clients sont les plus susceptibles de lui être infidèle et prendre les mesures appropriées pour lutter contre l'attrition. En outre, comme les clients ne partent pas tous pour les mêmes raisons, l'extraction de données permet aux fournisseurs de télécommunications sans fil de faire des offres et de proposer des promotions personnalisées aux clients concernés.

Cependant de nombreux analystes utilisent encore une approche traditionnelle d'extraction de données à l'aide de serveurs analytiques. Cette approche exploite les nombreux outils analytiques d'un serveur dédié qui permettent un haut niveau de flexibilité. Mais l'évolutivité face à un volume croissant de données n'est pas au rendez-vous. Ce manque d'évolutivité et de rapidité vient du fait que les modèles d'extraction de données ne peuvent pas tourner assez vite pour répondre aux besoins de l'entreprise, et par conséquent que les informations commerciales ne peuvent pas être mises à la disposition des représentants des services clients ou de la direction. Prenons l'exemple de notre hypothétique fournisseur de communications sans fil – le temps qu'il finisse d'analyser ses données historiques son client est déjà parti.

La capacité à extraire et à analyser les modèles qui se cachent sous des montagnes de données peut permettre aux organisations de mieux comprendre le comportement de leurs clients, les tendances de leurs marchés et donc de tirer profit de leurs informations. L'adoption d'une norme qui vient de faire son apparition le PMML ou Predictive Model Markup Language, sous XML, peut nettement faciliter les choses. Cette norme ouverte utilisée pour décrire des modèles statistiques et d'extraction de données, permet de définir puis de partager des modèles prédictifs entre applications. C'est le résultat du travail du Data Mining Group, un consortium d'éditeurs de logiciels fondé en 1998 afin de développer des normes d'extraction de données.

La charte PMML consiste à permettre à une application de produire un modèle et à une autre application d'utiliser ce modèle simplement en lisant le fichier de données PMML. Un modèle développé à l'aide d'un outil d'extraction peut ensuite être déployé ou exploité sur la totalité d'un entrepôt de données. Mais le PMML a ses faiblesses, surtout quand il s'agit de préparer des données brutes en vue d'analyse. Avec le PMML, les modèles ne sont pas appliqués directement sur les données détaillées de l'entrepôt de données, mais sur un jeu spécialisé de données sélectionnées pour leur valeur prédictive. Elles sont nettoyées ou transformées, prêtes à être traitées par les modèles analytiques. Le PMML permet certaines transformations de données pour des algorithmes spécifiques. Il est toutefois insuffisant quand il s'agit de gérer des processus complexes de nettoyage, de transformation et d'agrégation de toutes les données sélectionnées pour analyse. Le PMML part du principe que la préparation préalable des données a déjà été effectuée.

Cependant, si nous examinons le processus de modélisation analytique, il apparaît que plus de 70% du travail réside dans les étapes d'exploration et de préparation. Au cours de la phase d'exploration, l'analyste recherche les données présentant des structures répétitives ou des anomalies. Parmi les techniques les plus couramment utilisées, on compte les statistiques descriptives et la visualisation. L'exploration de données est itérative, c'est une tâche pour laquelle les données sont sollicitées de manière intensive et qu'il est préférable d'effectuer directement dans la base de données. Le prétraitement des données comprend la création du meilleur modèle analytique au moyen de la construction d'un jeu de données analytiques par la création de variables prédictives. L'analyste remplace les données manquantes ou non valides, et convertit les données sous un format adapté aux outils analytiques – par exemple, en remplaçant l'abréviation alphabétique d'un état par une valeur numérique pour le traitement. La phase de construction, de test et de validation de l'extraction de données est particulièrement intensive pour les unités centrales. Un environnement d'extraction de données ouvert permet aux analystes d'utiliser leurs outils d'extraction de données de prédilection, tout en effectuant les tâches lourdes de manipulation des données directement dans la base de données.

Dans sa version actuelle le PMML nécessite un coup de pouce. Celui-ci se concrétise sous la forme d'outils qui peuvent être utilisés pour construire les jeux de données analytiques. Ces outils ont la capacité d'explorer, de nettoyer, de transformer et d'agréger des données hautement normalisées au sein d'un entrepôt de données sous une forme acceptable par un éventail d'outils analytiques.

Des jeux de données analytiques permettent d'explorer et d'assembler les données requises et de les exploiter à l'aide des outils de modélisation aux normes de votre entreprise. Il s'agit de construire des modèles compatibles avec le PMML et de lancer le modèle dans la base de données en renvoyant le PMML à un consommateur de PMML. Ce qui signifie que les

activités de déploiement de modèles peuvent être effectuées directement à partir de la base de données plutôt que de devoir être transférées vers un serveur dédié.

Cette approche permet d'exécuter un modèle sur toutes les données et pas uniquement sur un sous-groupe. Par exemple, si le fournisseur de télécommunications sans fil mentionné plus tôt ne pouvait exécuter son modèle que sur un échantillon de sa clientèle, disons sur celle dont les contrats ont expiré, il ne pourrait pas voir les autres sous-groupes de sa clientèle potentiellement insatisfaits. S'il peut consulter la totalité de sa base de données, le prestataire de services peut repérer tous les secteurs de son activité menacés d'un taux d'attrition important quelle qu'en soit la raison.

Ajoutons également que le fait d'exécuter des modèles une seule fois par trimestre à cause du coût élevé de l'extraction de données est le meilleur moyen d'obtenir des informations erronées. En combinant les capacités du PMML avec un jeu de données analytiques, il est possible d'exécuter des modèles plus fréquemment et de continuer à exploiter la totalité du contenu d'un entrepôt de données. Au lieu de transférer les données vers les outils qui résident sur un serveur analytique, la méthode du jeu de données analytiques permet de gérer les fonctions et les traitements nécessaires directement dans la base de données. Les étapes d'exploration et de prétraitement des données, qui prenaient plusieurs semaines aux équipes d'extraction, peuvent maintenant se faire en quelques jours voire quelques heures.

## **Comment choisir un outil d'exploration de données**

Pour sélectionner un outil analytique beaucoup d'entreprise se contente de prendre l'outil que leurs spécialistes de l'exploration de données connaissent déjà. Cela peut paraître surprenant, mais comme la grande majorité des personnes qui ont suivi un troisième cycle ont dû utiliser des outils d'analyse de SAS Institute ou de SPSS Inc. cette voie est très souvent suivie, et cela confirme le principe que dans la nature c'est toujours la voie de la facilité qui prédomine. Cependant pour ceux qui ne veulent pas procéder ainsi les quatre critères suivants devraient faire partie de toute évaluation d'outils : profondeur de la technique d'analyse, largeur de la technique d'analyse, fonctions de traitement des données et facilité d'utilisation. (la profondeur fait référence à la qualité et à la fiabilité de chaque technique d'analyse, tandis que la largeur fait référence au nombre de techniques d'analyse prises en charge par l'outil.)

Il existe de nombreux candidats potentiels sur le marché, mais seule une douzaine d'entre eux font réellement ce qu'ils prétendent être capables de réaliser. L'évaluation de la qualité des algorithmes dissimulés sous les propos marketing requiert une grande expérience en statistique. Si vous ne disposez pas de cette expérience en interne, il est conseillé d'engager un consultant indépendant spécialisé dans l'analyse et l'exploration de données pour vous aider à évaluer et à réduire le choix des outils. Pour en revenir à vos besoins, un outil peut parfaitement convenir s'il traite une ou deux tâches d'analyse en profondeur, en complexité et deux ou trois autres superficiellement, pour autant que la profondeur de chaque technique corresponde à la profondeur de l'analyse nécessaire à la résolution de vos problèmes. Les outils d'analyse coûtent cher et il faut un certain temps avant d'apprendre à les utiliser efficacement. Il se peut qu'une entreprise sache qu'elle a besoin de tout un ensemble d'outils mais il est préférable de choisir au début un outil qui offre les principales fonctions requises, d'apprendre à s'en servir correctement et d'y ajouter par la suite des fonctions spéciales au fil de l'évolution des besoins en exploration de données.

Les deux plus grands soucis liés aux données portent sur le volume et la manipulation. Chaque outil possède une limite de volume. Il est important de voir si cette limite n'est pas trop basse. Si c'est le cas, cela signifie que l'outil a été conçu en tant que moteur d'analyse pour de petits volumes de données et non pour l'exploration de données. La possibilité de traiter de grands volumes est une exigence récente et de plus en plus présente. De nombreux outils sur le marché ont été développés avant l'émergence de ces nouvelles exigences au niveau des volumes et ils n'ont pas encore mis à jour leurs capacités. En plus du volume, la facilité des manipulations et des transformations des données est un facteur important à prendre en considération si vous souhaitez que la préparation des données soit efficace et bien gérée. On n'insistera jamais assez sur l'importance de découvrir et de tester les caractéristiques de chaque donnée, jusqu'à ce que l'on soit certain de posséder les bons éléments et que l'on connaisse toutes leurs caractéristiques. En règle générale, la qualité des résultats est directement proportionnelle au nombre d'itérations prises en charge par l'outil.

En matière de facilité d'utilisation, comme pour bien d'autres choses, plus un outil est complexe, plus l'apprentissage prendra du temps. Il se peut que l'outil avec l'interface utilisateur la plus simple et la plus intuitive permettent à vos spécialistes de l'exploration de données d'être opérationnels très rapidement, mais il peut également très vite atteindre ses limites face à la complexité croissante de l'exploration de données. Comme je l'ai déjà dit, la sélection de l'outil sophistiqué que vos spécialistes de l'exploration de données connaissent déjà a ses avantages. La phase d'apprentissage est raccourcie et il est fort probable qu'un tel outil offre une profondeur et une largeur bien plus grande qu'un outil à l'interface plus simple. Une autre stratégie consiste à établir un partenariat avec une personne qui connaît l'outil le plus complexe et qui sait comment réaliser des explorations de données jusqu'à ce que votre équipe interne ait acquis les compétences de base pour utiliser cet outil en toute confiance. C'est ce genre de transfert de compétence qui attire de nombreuses entreprises aux laboratoires d'analyse de données de sociétés de services spécialisées.

Alors que les éditeurs des outils qui font la course en tête continuent d'ajouter régulièrement des fonctions d'exploration de données et des techniques sophistiquées, l'exploration de données est de plus en plus efficacement utilisée dans le cadre de nombreux processus en particulier dans les domaines de la commercialisation et de la production. L'amélioration continue des outils permet une meilleure exploitation des entrepôts de données, si bien que l'exploration de données attire un nombre croissant d'adeptes. Et ce nombre croissant d'utilisateurs aux idées diverses va faire que l'exploration de données deviendra de plus en plus une discipline qui renforcera la qualité de la prise de décision en entreprise.

## **Propos sur l'analyse**

---

L'analyse est un processus structuré de construction, traitement et production d'informations. L'information, est à la fois matière première et produit généré. La pertinence de l'analyse dépend en partie de la qualité et parfois de la quantité des informations présent en compte. Mais par-dessus tout l'analyse nécessite une sérieuse dose de rigueur (démarche méthodique). Cependant il ne faut pas tomber dans un excès de rigueur, dans des analyses trop pointillistes interdisant une vue d'ensemble et amenant à ne pouvoir conclure, ou qu'avec des propositions précises mais très limitées, en refusant celles qui seraient globale mais imparfaitement vraies.

L'analyste ne doit pas se laisser submerger de données multiples, disparates, contradictoires parfois, et savoir identifier les informations cruciales, synthétiques, bien construites et suffisamment fiables. Cependant il faut avoir conscience que les données chiffrées et plus encore monétaires, même lorsqu'elles sont fausses avec précision - exemple comptabilité analytique avec critères de répartition obsolètes -, tendent à s'imposer aux données qualitatives, qui pourraient n'être qu'approximativement exactes. Les informations complètes, fines, certaines et rétrospectives sont généralement plus valorisées, que des informations partielles, agrégées, incertaines, tournées vers le futur.

Suivant leur (dé)formation d'origine, les analystes ont tendance à accorder trop de poids à certains facteurs. Les économistes privilégient les variables d'environnement au détriment de celles d'organisation ou de gestion. Les responsables des ressources humaines voient très systématiquement les causes fondamentales des difficultés, dans les jeux des acteurs, les blocages internes ou les qualifications des individus. Enfin les financiers cèdent régulièrement aux délices de l'analyse des flux de fonds, du bilan et du compte de résultat, pour y rechercher des déséquilibres qui ne font que traduire des dysfonctionnements stratégiques, organisationnels ou de gestion. Cependant, l'un des défis majeurs pour les analystes consiste à éviter les écueils dans les raisonnements : multiplication des perceptions, parcellisation des perceptions, éparpillement analytique, diversion-distraktion et approfondissement du faux problème.

S'il doit se méfier de lui-même, l'analyste doit faire preuve de prudence à l'égard des personnes et des groupes avec lesquels il est en relation. L'analyste peut se laisser prendre en otage par une direction qui cherche plus à faire entériner son point de vue qu'à obtenir des conclusions neutres. L'analyse peut être un alibi, une caution morale pour légitimer certaines décisions douloureuses. La lucidité de l'analyste, son aptitude à découvrir les buts cachés éventuels assignés à son intervention, sont essentiels à la réussite de sa mission. D'un point de vue déontologique, les conclusions d'un analyste n'ont pas à être inféodées aux intérêts de celui qui les a demandées, mais se doivent de servir en priorité l'entreprise dans son ensemble. L'analyste peut être confronté à des processus de désinformation, soit en provenance des sources extérieures, soit même au sein de l'entreprise. Sans tomber dans la paranoïa, il convient donc que l'analyste garde sa lucidité et fasse des recoupements.

Enfin, probablement on pourrait appliquer aux analystes en général, la boutade qui dit : « Si vous voulez cinq avis sur un problème économique, faites discuter quatre économistes ».

## Méthode et stratégie de résolution de problème

Pour résoudre un problème il est préférable de procéder avec méthode. Il existe diverses démarches plus ou moins élaborées de gestion des décisions comme par exemple les méthodes ESPRIT (Enoncé, Situation, Problématique, Résolution, Information, Terminaison) ou PORAS (Problème, Objectifs, Résolution, Articulation, Suite). Quoiqu'il en soit, au cœur de toutes ces approches se trouve l'information et dans le cas de problème de gestion dans l'entreprise d'aujourd'hui, le plus souvent le recours à des moyens informatiques dont certains spécifiquement conçus pour cela les entrepôts de données.

### Cycle de gestion d'une opportunité métier

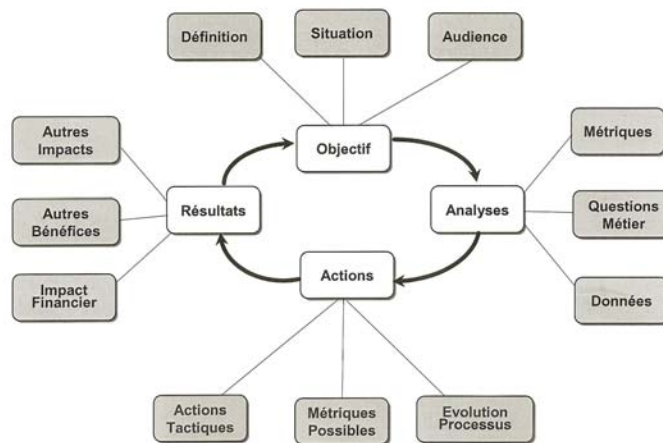


Illustration - 10

L'utilisation de données et leur transformation en information se fait à toutes les étapes de ces méthodes pour poser le problème à résoudre, pour l'analyser, pour déterminer des solutions et enfin pour envisager des décisions, leurs conséquences et les contrôles a posteriori possibles. Transformer des données en information est un savoir faire qui nécessite des compétences, une rigueur, cela implique de se situer par rapport à des références, de rajouter du spéculatif, du subjectif. Par exemple un problème n'existe pas en soi, mais pour soi, c'est une question de perception qui sera différente selon les personnes, les entreprises. Pour poser correctement un problème il convient donc de disposer d'informations sur la situation et ses différentes composantes (acteur, événement, milieu, perception, projection).

L'action de réunir, de traiter, d'organiser des informations relatives à un problème de gestion peut être facilitée par des moyens d'aide à la décision, mais quelle que soit l'aide apportée, les tâches à réaliser ne sont pas triviales. Il s'agit en effet de sélectionner les bonnes données, d'être à même de les analyser, de les relativiser, de déterminer des mesures, de les relier à des variables d'actions, de les partager avec les acteurs concernés. Sans parler des capacités de discernement de causes, de conséquences, de contradictions, ni celles de se référer à des principes ou des théories, ou de faire des hypothèses. Dans cet ensemble de tâches il convient en particulier d'avoir conscience que la plupart du temps il est réellement difficile de déterminer des mesures pertinentes.



Pour passer de l'information à la solution, il faut générer des solutions possibles, ce qui nécessite d'être à la fois rationnel et créatif. Ici encore il existe des méthodes comme celle de la recherche opérationnelle par exemple qui se fonde sur des modèles, des mesures et des critères d'optimisation ou toutes les méthodes de créativité qui favorisent la production d'idées. Il convient de prendre en compte le problème présent, à terme, les forces d'inertie, les tendances lourdes. Au final il s'agit de savoir combiner des variables d'actions (outil, mode opératoire, information, communication, formation).

Si les étapes évoquées ci-dessus peuvent être parfois compliquées, celle de la décision est très souvent complexe. Il s'agit de prendre en compte des critères de choix, des événements incertains, des interventions possibles d'autres acteurs, des conséquences attendues, des mesures préventives, de les scénariser et d'arrêter un choix. Ici encore il y a des méthodes, des théories qui peuvent aider notamment en fonction du type d'univers (déterminé, probabiliste, incertain, concurrentiel) dans lequel le choix doit être opéré. Par exemple la théorie des jeux modélise le comportement d'un agent face à des situations de choix dans lesquelles d'autres agents rationnels interagissent.

Au final une bonne résolution de problème se caractérisera par les qualités suivantes : économie de moyens, globalité (appréhension de l'ensemble du problème), dépassement de frontière, persévérance, préférence pour les approches connues. En termes de structure il convient pour les sujets lourds, comme par exemple la définition de l'évolution d'un système d'information, de mettre en place différents types de groupes spécialisés pour agir, pour informer, pour guider. Par exemple pour guider il faut s'appuyer sur des personnes différentes pour fixer des objectifs, définir des méthodes, contrôler les résultats.

Les sociétés de services ont l'habitude d'accompagner leurs clients dans la résolution de leurs problèmes de systèmes décisionnels, que ce soit pour rationaliser leurs moyens, pour y intégrer de nouvelles données ou pour exploiter sous de nouveaux angles les données historiques qu'ils ont déjà. Elles savent que dans les grandes entreprises il faut vouloir fortement les changements pour qu'ils arrivent et que vouloir est insuffisant. C'est pourquoi, outre les méthodes, le savoir faire de leurs consultants, l'organisation de contacts entre leurs clients, ils proposent des modèles de références qui permettent de se guider, de communiquer, de stimuler la créativité et globalement de gagner beaucoup de temps.

### **Analyse de données, information et stratégie : le modèle militaire !**

Faire de l'information une source de la compétitivité, c'est l'objectif que les fournisseurs de solutions décisionnelles proposent aux plus grandes entreprises mondiales à travers la mise en œuvre de ses solutions d'infrastructure de gestion et d'analyses de données. Les entrepôts de données facilitent la compréhension du passé et du présent, l'anticipation du futur, la réalisation des actions et la capitalisation des connaissances. Dans le contexte de mondialisation qui est le notre aujourd'hui, les entrepôts de données bouleversent nos habitudes de management, nos stratégies, nos besoins en compétences et nos capacités à nous adapter à un monde complexe en perpétuelle évolution. Il faut se rappeler la phrase de Darwin « Ce ne sont pas les espèces les plus fortes qui survivent, ni les plus intelligentes mais celles qui sont les plus à même de s'adapter aux changements ».

La première application des systèmes décisionnels a été l'aide aux décisions stratégiques. En effet, la stratégie qui était d'abord l'art de la guerre, étend désormais son application à la plupart des activités humaines et notamment aux entreprises. Information et stratégie sont étroitement liées car c'est à partir de la connaissance de la situation que l'action peut être décidée. Autrefois nous nous appuyions sur une logique en trois volets : stratégie/ mise en œuvre/ bilan. Aujourd'hui, la logique est plus cyclique : stratégie/ test/ réévaluation nécessaire/ aménagement de la stratégie/ test/ etc. car comme l'a dit un général américain lors de la dernière guerre d'Irak « aucun plan ne survit au contact de l'ennemi ».

L'information que nous prenons en compte dans le processus d'élaboration d'une stratégie est trop souvent l'information que nous avons nous-mêmes sélectionnée, c'est-à-dire celle qui peut servir notre stratégie. Pour ce motif, il est essentiel de séparer le donneur d'information du décideur. Il faut dans l'entreprise prendre comme modèle la sphère militaire, où l'on s'est toujours efforcé de séparer la responsabilité de l'information de celle de l'action. Sur un navire de guerre, par exemple, il existe d'un côté un officier spécifiquement chargé de recueillir et de traiter toutes les informations, et de l'autre des "metteurs en œuvre". Leurs fonctions respectives sont tout à fait exclusives, et cette dissociation est essentielle pour garantir une objectivité au niveau stratégique.

En stratégie, il faut appliquer des méthodes pour assurer une objectivité dans la prise de décision, mais force est de constater que les décideurs ont tendance à privilégier une perception intuitive des situations aux dépens des informations objectives. Dans les écoles de guerre on enseigne ainsi des méthodes qui prennent en compte la situation, les objectifs et les modes d'actions – des alliés comme des adversaires – et amènent à croiser ces données. L'exercice est souvent fastidieux mais très profitable. L'information est indispensable dans l'élaboration d'une stratégie, mais certaines données sont difficilement quantifiables, par exemple comment évaluer la résistance potentielle d'un concurrent ? Comment quantifier sa détermination ? De plus il faut se méfier de l'intox et les veilleurs (technique, marché, concurrence) doivent se rappeler la phrase de Sun Tzu « tout l'art de la guerre est fondé sur la duperie ». L'information est souvent un mélange de données quantitatives & qualitatives, et il est souvent difficile d'en obtenir une cohérence. C'est en fait le niveau de culture des co-décideurs et le dialogue qu'ils ont instaurés qui permettent cette cohérence.

Aujourd'hui, l'on tend à concentrer de nombreux systèmes au sein de systèmes d'information uniques, et de plus on tend à intégrer le décisionnel et l'opérationnel. Dans le monde militaire ce phénomène existe aussi depuis quelques temps, par exemple le porte-avions Charles de Gaulle, est un modèle du genre, il dispose déjà d'un système qui fonctionne en temps réel et qui est très structurant pour l'organisation du navire. En matière de commandement (management), tout ceci n'est pas sans inconvénient, car si l'on peut voir en temps réel depuis les quartiers généraux (directions générales) l'action sur le terrain, on prend le risque d'une trop grande ingérence des échelons supérieurs. Or le manque de connaissance des réalités du terrain (les systèmes appréhendent toujours partiellement la réalité) rend toute intervention de ce type dangereuse. Ceci explique le développement de la réflexion sur le « Pilotage Stratégique » dans les entreprises et la recherche de solutions (Balanced Scorecard, Business Activity Monitoring par exemple).

## **Décision, aide à la décision : un monde complexe**

Depuis quelques années, régulièrement des études sont réalisées sur la prise de décision dans les grandes entreprises et constatent que les interviewés disent devoir prendre de plus en plus de décision pour mener leur stratégie de gestion globalisée. Comme au même moment les responsables cherchent à être le plus rationnel possible et à tout aborder de façon scientifique, les fournisseurs de solutions se voient un bel avenir d'autant plus que pour le moment l'essentiel de leur activité se fait encore sur la mise en place de gisement de données, ce qui leur laisse un large domaine de développement dans la multiplication des applications de support aux décisions.

Le support qu'une entreprise attend d'un système d'information pour s'aider dans une prise de décisions est fonction du niveau et du type de décision concerné. Les gourous du management s'accordent pour distinguer quatre niveaux de décision pour les entreprises. Les décisions stratégiques définissent les relations de l'entreprise avec son environnement (choix des activités, des marchés, des filières technologiques), elles fixent des axes et des politiques pour le long terme. Les décisions d'organisation définissent les structures, les processus, les systèmes d'information (métier, mission, délégation, responsabilité, moyens) elles fixent le cadre de travail dans une perspective de moyen terme. Les décisions de gestion définissent pour le court terme la direction et la performance attendue (objectifs, ressources, contrôle). Enfin les décisions opérationnelles s'intéressent au fonctionnement quotidien tant au niveau de la supervision que de l'exécution.

Les décisions des différents niveaux peuvent être de différents types. Il y a les décisions non structurées, uniques, non routinières, ou l'incertitude est grande qui requiert beaucoup de jugement et présente des risques. A l'opposé il y a les décisions structurées, répétitives, routinières, aux procédures définies qui laissent peu de place à l'incertitude. Entre les deux il y a les décisions semi-structurées ou quelques facteurs introduisent des risques.

Face à une décision un responsable peut adopter différents comportements comme ne pas décider, s'en remettre au hasard, agir puis réfléchir, se fier à son intuition, réfléchir en solo, analyser des chiffres, se tourner vers un tiers, opter pour le collectif et enfin faire comme tout le monde. Qu'il travaille seul ou pas, en utilisant des moyens ou non, s'il ne s'abstient pas, il passe par quatre phases décisionnelles. Une phase d'instruction avec collecte d'information pour identifier le problème, les opportunités et définir ce qui peut faire une différence. Une phase de conception pour trouver, modéliser des options, définir et évaluer les solutions. Une phase de sélection de la ligne de conduite à adopter par comparaison des choix et prévisions des résultats. Enfin une phase de mise en application et d'évaluation des résultats.

Dans ce contexte rapidement esquissé et sans rentrer dans les dimensions fonctionnelles ou sectorielles, il est facile de comprendre que pour répondre à ces besoins il ne peut pas exister de solution universelle. Même le gisement de données détaillées que les fournisseurs de base de données affectionnent tant n'est pas l'unique source des informations de certains décideurs. Les entrepôts de données gèrent des informations « blanches » issues des activités historiques des compétiteurs d'un marché, mais dans certains cas il est nécessaire de s'appuyer sur des données « plus ou moins grises » issues de la veille, de l'intelligence économique ou du renseignement. Par exemple, dans le cadre d'une négociation le fait de disposer de renseignements précis sur les interlocuteurs clés peut être un atout décisif.

Il faut être humble, le domaine du décisionnel est particulièrement complexe beaucoup plus que celui de la gestion administrative des activités qui est couvert par les progiciels de

gestion intégrés par exemple. Une grande entreprise ne saurait couvrir ses besoins en s'appuyant sur les solutions d'un seul éditeur.

## **Au delà de la Décision, l'Action**

Le système d'information décisionnel des dirigeants et des collaborateurs d'une entreprise, a pour objet de permettre de rassembler et d'analyser des données concernant le passé, le présent et le futur, de façon à faciliter le suivi et la compréhension détaillée de l'environnement, des résultats, des acteurs internes ou externes, notamment les clients, pour optimiser les décisions stratégiques et la conduite des opérations.

Pour qu'une entreprise obtienne un avantage sur la concurrence, il convient qu'elle mobilise de nombreux acteurs dans le cycle qui mène de la donnée à l'action. En effet un avantage significatif s'obtient par la mise en œuvre d'actions permettant la réalisation profitable d'une opportunité : opportunité + action = avantage. L'opportunité elle-même est le produit d'une information et d'une décision : information + décision = opportunité. Dans ce cycle, différents acteurs (statisticiens, analystes métier, responsables d'activité, collaborateurs en charge des clients, des fournisseurs), sont impliqués aux différentes étapes qui voient la naissance des opportunités et leur réalisation.

Si la qualité d'un système d'information décisionnel permet d'espérer le succès, il est certain que celui-ci ne s'obtient pas de façon mécanique. Les paragraphes ci-dessous présentent de façon résumée, quelques idées à méditer sur le cycle : Données – Informations – Décisions – Actions.

Il est fondamental de bien séparer les données et l'information

Entre la donnée et l'information, au sens général du terme, il y a nécessairement un ou plusieurs médiateurs qui organisent, choisissent, rapprochent certaines données, décident d'en abandonner d'autres, pour pouvoir en tirer une sorte de synthèse d'un niveau plus élaboré que ses éléments constitutifs de base. Au final le travail de l'analyste est de dégager des faits, de les replacer dans une perspective, de les pondérer les uns par rapport aux autres, de les rendre clairs et compréhensibles : faits plus commentaires donnent alors l'information.

Transformer les données en information est un métier qui nécessite des compétences, une morale pour passer de la donnée à l'information, car cela implique d'y ajouter du subjectif, du spéculatif. Les informations utiles à la décision sont la plupart du temps extérieures à l'entreprise (clients, concurrence), et l'utilisateur doit cerner les informations dont il a besoin, or on entend souvent les utilisateurs se plaindre de ne pas avoir la "bonne" information. Souvent cela vient du fait que les définitions employées ne correspondent pas à leurs visions personnelles.

Une information n'a pour raison d'être que d'aboutir à une décision qui souvent se traduit en action

Les passages de la donnée à l'information et de l'information à la décision supposent une suite d'actions volontaires, réfléchies, intelligentes. Les mots information et décision sont tellement multiformes qu'ils peuvent avoir des sens contraires : une information peut être une désinformation, et une absence de décision faire office de décision. Décider est différent de

ratifier, c'est aussi créer et imaginer. Décision peut dire fixation de règles, étant entendu que l'application d'une règle ne suppose aucune prise de décision, mais par contre ne pas appliquer une règle est une décision. Souvent le contenu d'une décision sert d'information pour prendre une autre décision, on constate qu'il y a un réseau de décisions, avec des liens de dépendance et des liens pour information et pour action. Enfin, force est de constater que de nombreuses actions ne sont pas la suite d'un acte de décision.

La maîtrise de l'information est source de pouvoir

Les utilisateurs cherchent des systèmes à leur mesure et qu'ils maîtrisent. Ceci induit le risque de voir apparaître dans l'entreprise de multiples systèmes, non nécessairement cohérents entre eux, voire redondants. Les directions générales souhaitent des systèmes qui supportent en priorité leurs initiatives stratégiques et assurent la cohérence d'ensemble de l'entreprise. Ceci conduit à mettre en place des systèmes qui visent à partager les informations entre les différents utilisateurs.

Les informaticiens veulent bâtir des systèmes qui répondent à tous les besoins, qu'ils soient exprimés ou non. L'intention est louable, mais certainement un peu utopique, car les activités et les besoins d'une entreprise sont par essence évolutifs. Cependant suivant l'infrastructure mise en place (systèmes indépendants ou système d'entreprise), il est plus ou moins facile de concilier personnalisation, performance, cohérence et évolution.