



## ANALYSE PRÉDICTIVE DU BIG DATA

Comment bénéficier d'un avantage concurrentiel en combinant le Big Data à l'analyse prédictive

# TABLE DES MATIÈRES

RETOUR VERS LE FUTUR : L'ÉMERGENCE DU PHÉNOMÈNE BIG DATA	1
EXPLOITER L'OPPORTUNITÉ COMMERCIALE D'« ANALYSE PRÉDICTIVE DU BIG DATA »	1
POINT N° 1 : PLATE-FORME INFORMATIQUE AGILE	2
MISE EN PLACE D'UN ENTREPÔT DÉCISIONNEL « AGILE »	2
ENTREPÔT DÉCISIONNEL ET ANALYSE PRÉDICTIVE INTÉGRÉS	2
POINT N° 2 : ÉVOLUTIVITÉ LINÉAIRE	3
CONVERSION DE L'ETL EN PROCESSUS D'ENRICHISSEMENT DES DONNÉES	3
GESTION DE CHARGES D'ANALYSE ET DE REQUÊTE EXTRÊMEMENT VARIABLES	3
ANALYSE DE DATASETS GRANULAIRES VOLUMINEUX (BIG DATA)	4
ACCÈS AUX DONNÉES ET PRISE DE DÉCISION À FAIBLE TEMPS DE LATENCE	4
POINT N° 3 : ENVIRONNEMENT UTILISATEUR OMNIPRÉSENT, UNIVERSEL ET COLLABORATIF	5
MISE EN OEUVRE D'UN ENVIRONNEMENT UTILISATEUR INTUITIF ET OMNIPRÉSENT	5
EXPLOITATION DE L'ASPECT COLLABORATIF	5
MISE EN OEUVRE DE NOUVELLES APPLICATIONS MÉTIERS	5
APPLICATIONS BASÉES SUR DES ATTRIBUTIONS	5
APPLICATIONS BASÉES SUR DES RECOMMANDATIONS	6
APPLICATIONS PRÉDICTIVES/PRÉVISIONNELLES	6
APPLICATIONS D'ANALYSE DÉTAILLÉE	7
APPLICATIONS D'ANALYSE COMPARATIVE	7
CONCLUSION	7
À PROPOS DE L'AUTEUR	8

## RETOUR VERS LE FUTUR : L'ÉMERGENCE DU PHÉNOMÈNE BIG DATA

Au cours des années 1970 et au début des années 1980, les fabricants et les revendeurs de biens de grande consommation (BGC) ont dirigé leurs activités en s'appuyant sur les études de marché bimestrielles du panel d'AC Nielsen. Ces études présentaient des données relatives à la concurrence et au marché (chiffre d'affaires, unités vendues, prix moyen, part de marché, par exemple). Les fabricants de BGC les exploitaient pour définir leurs dépenses, plans et stratégies (marketing, commerciales, publicitaires et promotionnelles) avec leurs réseaux de partenaires (distributeurs, grossistes, revendeurs, etc.).

Par la suite, au milieu des années 1980, l'entreprise Information Resources Inc. (IRI) a commencé à proposer l'installation de scanners de point de vente (PDV) gratuits dans les boutiques en échange des données de scanner résiduelles. Les revendeurs se réjouirent de cette proposition, sachant que la main-d'oeuvre constituait leur principal poste de dépense et que la valeur des données de scanner ne leur paraissait pas évidente.

Ces données de scanner PDV, alors considérées comme du « Big Data », changèrent les règles du jeu, provoquèrent un transfert de pouvoir dans le secteur (entre les fabricants et les revendeurs) et modifièrent la manière dont les deux parties géraient leurs activités. Les mégaoctets de données cédèrent rapidement la place à des téraoctets, qui nécessitèrent l'utilisation d'une nouvelle génération de plates-formes de stockage et de serveurs (par ex., Teradata, Red Brick, Sybase IQ, Sun et Sequent) et d'outils d'analyse (tels que Metaphor, Business Objects, Informatica et MicroStrategy).

Les entreprises à la pointe du progrès, telles que Wal-Mart, Procter & Gamble, Tesco et Frito Lay, exploitèrent ce nouveau « Big Data », ainsi que les outils et plates-formes d'analyse novateurs afin de se démarquer de leurs concurrents. Ces entreprises se trouvaient au premier plan du développement de nouvelles catégories d'applications métiers orientées vers l'analyse prédictive et le Big Data pour trouver une solution aux problèmes qui, auparavant, étaient impossibles à résoudre de manière économique (voir l'encadré ci-contre).

Trente ans plus tard, tout recommence. Nous assistons à l'explosion de nouvelles sources de données diverses à granularité fine et à faible latence (dites « Big Data »), capables de révolutionner le mode de fonctionnement des entreprises et des industries. Ces nouvelles sources de données proviennent d'un ensemble de périphériques, d'interactions client et d'activités métiers qui mettent en lumière de nouvelles analyses des chaînes de valeur organisationnelle et sectorielle.

L'émergence de ces nouvelles sources de données, plus détaillées, permet aux entreprises de saisir des opportunités commerciales qui leur échappaient jusque-là et de créer de nouvelles gammes d'applications métiers. Cependant, pour que cela devienne réalité, il faut de nouvelles plates-formes (une infrastructure) et de nouveaux outils (l'analyse prédictive). Ce livre blanc souligne le rôle potentiel de ces nouvelles plates-formes et évoque les types d'application métier axés sur l'analyse prédictive et le Big Data qui pourraient émerger.

## EXPLOITER L'OPPORTUNITÉ COMMERCIALE D'« ANALYSE PRÉDICTIVE DU BIG DATA »

Le Big Data nécessite une nouvelle plate-forme d'analyse prédictive qui permettra aux entreprises comme aux technologies de bénéficier d'un avantage concurrentiel. Pour cela, il convient d'utiliser une nouvelle infrastructure technologique (a) à même d'évoluer vers des pétaoctets de données, (b) capable de prendre en charge l'accès aux données à faible latence et la prise de décision et (c) offrant des fonctions d'analyse prédictive intégrées pour accélérer les processus sophistiqués de mise en oeuvre et de modélisation analytiques.

En démultipliant la puissance à sa disposition pour traiter les Datasets volumineux, l'entreprise est alors en mesure d'identifier en continu des données exploitables enfouies dans le Big Data et d'intégrer de manière transparente ces données dans l'environnement de travail de l'utilisateur dès que cela se révèle opportun. Cette nouvelle plate-forme d'analyse prédictive libère les entreprises des vieilles méthodes de reporting rétrospectif en leur fournissant une analyse prédictive et prospective des énormes volumes de données et en favorisant la prise de décisions plus pertinentes à tous les niveaux de l'entreprise.

### Nouvelles catégories d'applications métiers basées sur l'analyse prédictive :

- Préviation en fonction de la demande
- Optimisation de la chaîne d'approvisionnement
- Efficacité des dépenses en prospection commerciale
- Analyse du panier de la ménagère
- Gestion des catégories et merchandising
- Optimisation des prix/du rendement
- Gestion de la démarque des produits
- Programme de fidélisation de la clientèle

## POINT N° 1 : PLATE-FORME INFORMATIQUE AGILE

L'agilité repose sur des architectures d'entrepôt décisionnel et d'analyse prédictive extrêmement flexibles et reconfigurables. Il est possible de reconfigurer et de redéployer rapidement les ressources d'analyse en vue de répondre à l'évolution constante des exigences de l'entreprise et d'atteindre ainsi des niveaux élevés de flexibilité et d'agilité.

### MISE EN PLACE D'UN ENTREPÔT DÉCISIONNEL « AGILE »

La nouvelle plate-forme d'analyse prédictive permet de développer des entrepôts décisionnels libérés des contraintes des environnements informatiques actuels. De nos jours, les entreprises sont obligées d'employer des techniques de conception insolites et des outils de reporting rudimentaires pour générer des analyses détaillées des sources de données à croissance exponentielle en recourant à des technologies de base de données obsolètes. Alors que ces volumes de données ne cessent d'augmenter et que de nouvelles sources de données apparaissent en ligne, les entreprises s'aperçoivent que les architectures, les outils et les solutions actuels sont trop onéreux, trop lents et trop rigides par rapport à leurs initiatives métiers stratégiques.

Il est nécessaire d'envisager l'impact des agrégats prédéfinis. Les agrégats<sup>1</sup> servent généralement à compenser la puissance de traitement limitée des systèmes traditionnels de gestion de base de données relationnelle dans le cadre du traitement des jointures multitable et des analyses de tables massives. Un administrateur de base de données précalcule les agrégats les plus courants lors de la préparation des données afin d'améliorer les performances ad hoc et de reporting. Le volume des données stockées dans ces tables d'agrégats est plusieurs fois supérieur à celui des données brutes elles-mêmes. Le précalcul des agrégats dure si longtemps que les contrats de niveau de service en pâtissent. L'exploitation du « chargement en flux continu » destiné à générer des « rapports opérationnels en temps réel » est une pure fiction compte tenu du temps nécessaire à la reconstruction des tables d'agrégats lors de l'entrée de nouveaux « flux » de données dans l'entrepôt décisionnel.

L'élimination de ces limitations permet de mettre en oeuvre un environnement d'entrepôt décisionnel agile aussi flexible et réactif que l'entreprise qui l'utilise en exploitant les fonctions suivantes :

- **Agrégation à la demande** : il est inutile de prédéfinir des agrégats pour réduire le temps de réponse des requêtes et des rapports. Il est possible de créer des agrégats en temps réel, ce qui élimine la nécessité plutôt décourageante de reconstruire en permanence des agrégats chaque fois que de nouveaux flux de données entrent dans l'entrepôt décisionnel.
- **Indépendance par rapport à l'indexation** : les administrateurs de base de données peuvent se passer de l'indexation rigide. Ils n'ont pas besoin de connaître à l'avance les questions que les utilisateurs poseront de manière à créer tous les index sous-jacents. Les utilisateurs sont libres de demander le niveau supérieur de questions métiers détaillées sans se préoccuper des problèmes de performances.
- **Création directe d'indicateurs clés de performance** : les professionnels sont libres de définir, de créer et de tester les nouveaux indicateurs clés de performance dérivés (et composites) sans avoir à impliquer les administrateurs de base de données pour les précalculer.
- **Structures hiérarchiques ad hoc flexibles** : il est inutile de prédéfinir les hiérarchies dimensionnelles lors de la création de l'entrepôt décisionnel. Par exemple, au cours d'une analyse de marché, les entreprises sont en mesure, si elles le souhaitent, de changer les entreprises auxquelles elles sont comparées.

### ENTREPÔT DÉCISIONNEL ET ANALYSE PRÉDICTIONNELLE INTÉGRÉS

Traditionnellement, l'entrepôt décisionnel et l'analyse prédictive ont toujours résidé dans des environnements différents. Le déplacement des données de l'entrepôt décisionnel vers l'environnement d'analyse prédictive exige un processus ETL (d'extraction, de transformation et de chargement) distinct au sein duquel les données sont sélectionnées, filtrées, agrégées, prétraitées, reformatées, puis transportées vers l'environnement d'analyse prédictive. Dès lors que les données sont stockées dans l'environnement d'analyse prédictive, les analystes créent, testent et affinent les modèles et les algorithmes analytiques. Si, au cours du processus, les analystes se rendent compte qu'ils ont besoin de données plus granulaires et/ou de données différentes, ils doivent recommencer l'ensemble du processus ETL de l'entrepôt décisionnel, ce qui peut rallonger le processus analytique de plusieurs jours, voire de plusieurs semaines.

<sup>1</sup> Les agrégats sont des résumés précalculés hiérarchiques ou dimensionnels des faits (mesures ou audits), généralement définis par la clause « SQL Group By ». Dans une dimension géographique, par exemple, vous pouvez créer des agrégats pour l'ensemble des faits (ventes, chiffre d'affaires, marge exprimée en euros, marge définie en pourcentage, rendement, etc.) par pays, région, province, département, ville et code postal.

Un environnement d'entrepôt décisionnel et d'analyse prédictive intégré unique (avec analyse prédictive au sein de la base de données) signifie que les analystes peuvent demeurer sur le site de l'entrepôt décisionnel pour effectuer les opérations d'analyse prédictive. Les Datasets volumineux peuvent être transférés entre les environnements d'entrepôt décisionnel et d'analyse prédictive à des vitesses extrêmement élevées (de 5 To à 10 To/h). Conséquence : un processus d'analyse considérablement accéléré et une intégration plus facile des résultats d'analyse prédictive dans l'environnement d'entrepôt décisionnel et de Business Intelligence. Par exemple, un environnement d'entrepôt décisionnel et d'analyse prédictive intégré prend en charge les types d'analyse suivants :

- Sous-segmentation et transmission des Datasets volumineux entre l'entrepôt décisionnel et l'environnement d'analyse prédictive en vue de créer des « sandbox analytiques » pour l'exploration et la découverte analytiques.
- Interrogation des Datasets volumineux au niveau de granularité le plus bas pour signaler les tendances, les activités et les comportements « inhabituels » afin de générer des informations exploitables assorties des recommandations correspondantes.
- Accélération du développement et des tests des différents scénarios métiers afin de faciliter les analyses par hypothèses, de sensibilité et de risques.

Ces types d'avantage seraient inestimables s'ils étaient intégrés dans les tâches quotidiennes. Il suffit d'aborder ce sujet avec un responsable de la distribution qui souhaite que le système signale les problèmes potentiels liés aux performances des vendeurs concernant une dégradation du service, mesurée par rapport au respect des délais de livraison et du pourcentage des livraisons complètes réalisées.

## POINT N° 2 : ÉVOLUTIVITÉ LINÉAIRE

L'accès à une puissance de traitement massive implique que les problèmes métiers soient envisagés sous un angle radicalement différent. Passons en revue quelques exemples dans lesquels l'évolutivité informatique massive peut avoir un impact sur l'entreprise.

### CONVERSION DE L'ETL EN UN PROCESSUS D'ENRICHISSEMENT DES DONNÉES

L'ETL vise à corriger les erreurs générées par les systèmes sources en extrayant, en convertissant, en nettoyant, en profilant, en normalisant et en alignant l'ensemble des données afin de garantir que les utilisateurs comparent effectivement des éléments analogues dans leurs analyses. Grâce à la puissance de traitement disponible pour l'ETL (en plus des nouveaux langages informatiques tels que Hadoop), il est possible de convertir le processus ETL classique en processus d'enrichissement des données. Vous avez ainsi la possibilité de créer des audits nouveaux et utiles, notamment les suivants :

- **Séquencement ou classement des activités** : Identification de la séquence d'activités exécutées avant un événement donné. Par exemple, identification d'une personne qui appelle systématiquement le centre d'appels deux fois après avoir recherché des options de support sur le site Web avant que le problème ne soit résolu.
- **Fréquence** : Nombre d'occurrences d'un événement donné pendant une période donnée. Par exemple, identification des produits ayant fait l'objet de x appels au centre de support au cours des 90 premiers jours d'utilisation.
- **N-tiles** : Regroupement d'éléments (produits, événements, clients et partenaires) sous forme de compartiments en fonction d'un audit ou d'un groupe d'audits. Par exemple, suivi du tertile supérieur (les premiers 10 %) des clients en fonction du chiffre d'affaires ou de la marge sur une période de trois mois consécutifs.
- **« Paniers » comportementaux** : Création d'un « panier » d'activités, comprenant la fréquence et le classement, qui précèdent une vente ou une « conversion » afin d'identifier les combinaisons de traitements commerciaux les plus efficaces et les plus rentables.

### GESTION DE CHARGES D'ANALYSE ET DE REQUÊTE EXTRÊMEMENT VARIABLES

Il est très difficile de prévoir les types de requête et d'analyse prédictive que l'entreprise pourrait souhaiter exécuter en fonction de l'environnement métier le plus récent. Les actions sur les prix ou les actions promotionnelles d'un concurrent peuvent nécessiter de procéder immédiatement à une analyse afin de mieux comprendre l'impact financier et économique sur l'entreprise. Les analyses prédictives les plus pertinentes sont celles qui impliquent des charges de travail extrêmement variables, difficilement prévisibles a priori.

Auparavant, les entreprises devaient se contenter d'analyses approximatives « après coup ». Elles ne disposaient pas non plus de la puissance de traitement nécessaire pour effectuer une analyse approfondie des événements « à chaud » ou pour tenir compte de l'ensemble des paramètres et changements susceptibles d'avoir une incidence sur leur activité. Or, les nouvelles plates-formes sont capables de satisfaire à ces besoins d'analyses instantanées, extrêmement exigeantes en puissance de traitement. Cette fonction se présente de la manière suivante pour les professionnels :

- **Performances et évolutivité** : agilité requise pour analyser de manière détaillée les données afin de poser les questions de deuxième et de troisième niveaux nécessaires à la prise de décision. Si les professionnels souhaitent étudier les données dans le moindre détail afin d'identifier les variables conditionnant l'activité de l'entreprise, ils n'ont pas à s'inquiéter du risque de paralysie éventuelle du système suite à l'analyse d'une trop grande quantité de données.
- **Agilité** : prise en charge du développement rapide, des tests et de l'affinement des modèles analytiques permettant de prévoir les performances de l'entreprise. Les analystes des données peuvent décider d'explorer les différentes variables pouvant avoir un impact sur les performances de l'entreprise, tirer des conclusions à partir des résultats et intégrer ces conclusions dans l'itération suivante du modèle. Ils sont libres d'arrêter l'analyse au premier échec sans se soucier des effets sur les performances du système.

#### ANALYSE DE DATASETS GRANULAIRES VOLUMINEUX (BIG DATA)

L'un des avantages les plus significatifs du cloud est la quantité et le niveau de données détaillées pouvant être exploités lors de l'analyse et de la modélisation des facteurs décisionnels de l'entreprise. Le cloud fournit non seulement le potentiel d'une puissance de traitement à la demande plus efficace, mais également des fonctions de stockage de données plus économiques et efficaces. Au lieu d'être contrainte à s'adapter aux données, l'entreprise est libre d'étendre son analyse en tirant parti de tous les aspects des données des manières suivantes :

- Possibilité d'exécuter des analyses multidimensionnelles jusqu'au nième degré. L'entreprise ne doit pas nécessairement se limiter à envisager trois ou quatre dimensions ; elle peut en étudier des centaines, voire des milliers, en vue d'affiner et d'identifier ses performances. Grâce à ce niveau d'analyse multidimensionnelle, elle peut déterminer les moteurs d'activité en fonction d'une donnée géographique spécifique (ville ou code postal, par exemple), du produit (numéro SKU ou code UPC), du fabricant, de la promotion, du prix, de l'heure ou du jour de la semaine, etc. Avec ce niveau de granularité, il est possible d'améliorer de façon très nette les performances métiers locales.
- Recherche d'un nombre suffisant de « petites » pépites dans le volume de données, capables de représenter une différence tangible pour l'entreprise. La plate-forme relève deux principaux défis de l'analyse localisée : rechercher les facteurs décisionnels au niveau local ou à un niveau spécifique et rechercher un nombre suffisant de ces facteurs locaux qui peuvent créer une différence tangible pour l'entreprise.

#### ACCÈS AUX DONNÉES ET PRISE DE DÉCISION À FAIBLE TEMPS DE LATENCE

Comme les données ne passent pas par une longue étape de préparation (par rapport aux agrégats prédéfinis et aux mesures dérivées précalculées), le temps de latence entre la génération des données et la mise à disposition de ces dernières pour l'entreprise est considérablement réduit. La possibilité de réduire le laps de temps entre ces deux étapes est l'incarnation du concept d'analyse opérationnelle et ce, des manières suivantes :

- Exploitation en continu des flux de données (chargements en flux continu) en vue de générer des fonctions de reporting et d'analyse prédictive opérationnelles à faible latence. Le délai entre un événement d'entreprise (tel qu'une valeur mobilière) et la décision d'achat ou de vente s'en trouve considérablement réduit. Nous pouvons clairement observer l'impact de cette prise de décision à faible latence avec l'essor des opérations algorithmiques de Wall Street<sup>3</sup>.
- L'accès à faible latence aux données permet une prise de décision rapide en temps opportun. Ainsi, les gestionnaires de campagnes peuvent réallouer les budgets de campagne en ligne aux sites et aux combinaisons de mots clés les plus efficaces et/ou assurant la meilleure conversion alors qu'une campagne est en cours.

## POINT N° 3 : ENVIRONNEMENT UTILISATEUR OMNIPRÉSENT, UNIVERSEL ET COLLABORATIF

Les professionnels ne demandent pas plus de données, de graphiques et d'options de reporting, aussi esthétique que soit leur présentation. Ils veulent une solution qui tire le meilleur parti de l'analyse prédictive afin d'identifier et d'obtenir des informations exploitables, matérielles, sur l'entreprise.

### MISE EN OEUVRE D'UN ENVIRONNEMENT UTILISATEUR INTUITIF ET OMNIPRÉSENT

Le couplage des données détaillées et de la puissance de l'analyse massive présente un avantage intéressant : la création d'interfaces plus simples et plus intuitives. Comment est-ce possible ? Pensez à la relation entre l'iPod et iTunes. L'interface minimaliste de l'iPod explique en partie son succès commercial (et sa prédominance sur le marché). Apple a transféré la complexité d'utilisation (par ex., gestion des listes de lecture, ajout de chansons et génération de recommandations en utilisant la fonction Genius) de l'iPod vers iTunes où elle est plus simple à gérer. Nous pouvons appliquer ce concept en vue d'améliorer l'environnement utilisateur de l'analyse prédictive.

- L'environnement utilisateur permet de tirer le meilleur parti de l'analyse prédictive afin de s'acquitter d'une plus grande partie des lourdes tâches d'analyse exécutées en arrière-plan. Au lieu de présenter des rapports, des graphiques et des feuilles de calcul toujours plus complexes, l'interface devient plus intuitive et fournit aux utilisateurs les informations détaillées dont ils ont besoin concernant l'activité de l'entreprise.
- En s'appuyant sur les informations détaillées extraites des données, l'environnement utilisateur peut entraîner des actions recommandées précises (à l'instar de la fonction Genius d'iTunes). La complexité de l'identification des recommandations pertinentes et exécutables incombe à l'analyse prédictive.

Imaginez, par exemple, l'interface d'un responsable de campagne publicitaire qui limite la myriade de variables liées aux performances des campagnes aux seules variables matérielles et exploitables. Supposez encore que l'interface utilisateur présente non seulement ces variables, mais aussi des recommandations destinées à améliorer les performances des campagnes en cours. Il s'agit effectivement d'un environnement utilisateur dont rêvent tous les professionnels.

### EXPLOITATION DE L'ASPECT COLLABORATIF

La collaboration est une composante naturelle des processus d'analyse et décisionnels. De petites communautés d'utilisateurs partageant des caractéristiques communes se constituent rapidement pour échanger leurs connaissances sur des sujets spécifiques.

Il serait très utile, par exemple, que tous les directeurs de marque d'une grande entreprise de produits de grande consommation créent une communauté dans laquelle ils pourraient partager des données, informations et visions relatives à la gestion des marques et en discuter. Les campagnes marketing concluantes pour une marque pourraient alors être copiées plus rapidement par d'autres marques et être étendues via le partage des données et des analyses résultantes.

### MISE EN OEUVRE DE NOUVELLES APPLICATIONS MÉTIERS

L'une des manières d'apprécier le potentiel de cette nouvelle plate-forme d'analyse prédictive consiste à se poser la question suivante : avec sa puissance de traitement à la demande, ses Datasets à granularité fine, son accès à faible latence aux données et l'intégration étroite de l'entrepôt décisionnel et de l'analyse prédictive, quels types de problème métier que l'entreprise ne pouvait pas résoudre auparavant est-elle à présent capable d'appréhender ? Examinons quelques applications métiers qu'autorise l'analyse prédictive sur ces nouvelles plates-formes, notamment en combinaison avec le Big Data.

### APPLICATIONS BASÉES SUR DES ATTRIBUTIONS

Les applications basées sur des attributions sont destinées à attribuer un « crédit » concernant un événement donné à une série d'activités ou de transactions au sein d'un processus complexe, composé de plusieurs étapes. Ces applications doivent récupérer, aligner et analyser la série d'activités en tenant compte de facteurs tels que la fréquence, le séquençement, le caractère récent, les seuils et le déclin entre les activités, afin de créditer une valeur à chaque activité. Exemples d'applications basées sur des attributions :

<sup>2</sup> Sur les marchés financiers électroniques, le commerce algorithmique consiste à utiliser des programmes informatiques pour passer des ordres, l'algorithme informatique se chargeant de déterminer divers aspects de l'ordre, tel que le moment, le prix ou la quantité ou de lancer l'ordre sans aucune intervention humaine.

- Applications d'efficacité marketing multicanal dans le cadre desquelles les services commerciaux tentent d'attribuer un crédit relatif à une vente sur plusieurs canaux marketing. Cela s'applique particulièrement aux distributeurs en ligne qui s'efforcent d'attribuer un crédit pour une conversion dans plusieurs affichages publicitaires, sites Web et recherches de mots clés.
- Applications d'attribution orientées partenaires dans le cadre desquelles les services commerciaux tentent de mesurer les contributions des partenaires dans les transactions métiers complexes en plusieurs étapes.
- Applications d'attribution orientées traitements médicaux dans le cadre desquelles les établissements de santé tentent d'attribuer l'impact de différents traitements et médicaments ayant eu un certain effet.

#### APPLICATIONS BASÉES SUR DES RECOMMANDATIONS

Les applications basées sur des recommandations identifient et créent des groupes d'utilisateurs ou de produits semblables ou similaires en fonction des comportements, des données démographiques ou d'autres attributs discernables. Les applications analysent les transactions à partir de ces groupes pour créer des propensions chargées de mesurer la solidité des relations entre les utilisateurs et leurs comportements et préférences. À partir de ces propensions, les applications peuvent émettre des recommandations de produits (Amazon et Netflix, par exemple) ou de personnes (LinkedIn et Facebook, par exemple). Exemples d'applications basées sur les recommandations :

- Applications de ciblage publicitaire des clients qui recommandent des segments d'audience cibles « semblables » ou similaires en fonction de comportements et d'un historique d'achat de produits (par exemple, il est fort probable que les campagnes qui touchent avec succès les « Soccer Moms » aux États-Unis auront également le même impact sur les « New Grannies »).
- Applications qui recommandent des produits complémentaires en fonction de ce que des utilisateurs similaires ont acheté à un moment donné (par exemple, les clients qui ont acheté une nouvelle habitation dans des régions correspondant à certains codes postaux sont susceptibles d'acheter une nouvelle machine à laver et un nouveau sèche-linge dans les trois mois suivant l'achat de leur logement).

#### APPLICATIONS PRÉDICTIVES/PRÉVISIONNELLES

Les applications prédictives et prévisionnelles acquièrent un large éventail de variables, d'audits et de dimensions afin de faciliter la prise de décision dans différents scénarios commerciaux. Ces applications tirent le meilleur parti des techniques de statistique et d'analyse des données afin d'exploiter une multitude de variables destinées à identifier les facteurs (et combinaisons de facteurs) les plus efficaces pour prévoir les performances dans des cas précis. Étant donné l'horizon temporel de certaines de ces décisions (telles que la fixation des prix), l'accès à faible latence aux données et l'analyse prédictive intégrée dans la base de données sont des éléments déterminants dans la réussite de l'opération.

Les applications prédictives avancées reposent sur l'évaluation des risques et de la sensibilité, ce qui permet aux décideurs d'identifier les variables les plus importantes dans la prise de décision. Si, par exemple, une variable se révèle critique dans la prise de décision, il est alors possible d'en vérifier l'exactitude et la globalité. Exemples d'applications prédictives/prévisionnelles :

- Applications de détermination de l'attrition des clients qui prévoient la probabilité d'attrition des clients en fonction de critères, tels que les activités d'utilisation, les demandes de support, les modèles de paiement et l'influence sociale des amis.
- Applications de maintenance des produits qui prévoient les défaillances d'équipement en fonction des informations d'utilisation du produit (notamment celles dorénavant fournies par les périphériques de données intégrés), des records de service de maintenance et l'historique général des performances du produit.
- Applications de mesure des performances des employés qui prévoient les performances potentielles d'un employé en fonction de critères, tels que la formation, la classe socio-économique, l'historique des emplois précédents, l'état civil et des réponses psycho-comportementales.
- Applications de test clinique qui modélisent l'effet de différents médicaments en fonction de tests cliniques afin que l'entreprise puisse comprendre l'efficacité de certains traitements et éviter les conséquences catastrophiques de l'utilisation de certaines associations médicamenteuses. Cela est d'autant plus important lorsqu'il s'agit de déterminer les effets relatifs à plusieurs traitements et médicaments (voir les applications basées sur les attributions).

- Applications de gestion du rendement, de la démarque et d'optimisation des prix destinées à créer des modèles temporels afin de permettre aux décideurs de déterminer quand et de combien il convient d'augmenter ou de réduire les prix en fonction des conditions actuelles de l'offre et de la demande. Ces types d'application sont largement employés dans le secteur des biens de consommation (denrées périssables, billets d'avion, chambres d'hôtel, vêtements à la mode et billets d'événements sportifs, par exemple) dont la valeur atteint zéro à un moment donné.

#### APPLICATIONS D'ANALYSE DÉTAILLÉE

Les applications d'analyse détaillée utilisent des techniques de calcul de statistiques et d'analyse des données pour identifier les situations ou les comportements « inhabituels ». Les applications avancées d'analyse détaillée permettent de réaliser des analyses complexes aussi bien sur des centaines que des milliers de dimensions métiers. Ces applications jouent un rôle prépondérant face à l'augmentation du volume de données provenant de sources telles que les clics Web, les détecteurs RFID et les appliances en réseau. Exemples d'applications d'analyse détaillée :

- Applications de distribution des produits et de démarque qui surveillent constamment des détecteurs et les données RFID afin d'identifier les différences entre l'emplacement supposé et l'emplacement réel d'un produit.
- Applications antifraude destinées à surveiller en permanence les transactions financières afin d'identifier les comportements « inhabituels » pouvant révéler des activités frauduleuses. Ces types d'application conviennent particulièrement aux cartes de crédit, aux comptes bancaires et aux remboursements de frais médicaux.
- Applications de blanchiment d'argent sale destinées à surveiller en permanence les flux d'argent pour identifier les situations « inhabituelles », telles qu'un nombre anormalement élevé de transactions séquentielles d'un petit montant, signe potentiel d'un blanchiment d'argent.

#### APPLICATIONS D'ANALYSE COMPARATIVE

Les applications d'analyse comparative tirent le meilleur parti d'analyses prédictives afin de comparer les performances d'une entité à une référence. La référence de comparaison peut correspondre à une norme sectorielle, à une période ou à un événement passé (campagne marketing, par exemple). Exemples d'applications d'analyse comparative :

- Applications de part de marché conçues pour présenter des informations de part de marché et de part de portefeuille. Par exemple, les grandes entreprises présentes sur le Web peuvent fournir des données et des analyses de « part de publicité » qui permettent aux annonceurs et aux agences de connaître leurs investissements publicitaires par rapport à ceux de la concurrence.
- Applications d'analyse comparative de la concurrence qui comparent les performances d'une entreprise à un agrégat de concurrents ou à une moyenne sectorielle. Les entreprises disposent ainsi d'une référence par rapport à laquelle elles peuvent comparer leurs résultats financiers ou commerciaux.
- Applications d'analyse comparative de campagne qui mesurent les performances d'une campagne marketing en cours à une campagne ou à un événement marketing antérieur. Par exemple, une entreprise peut vouloir comparer les performances de sa campagne actuelle du « 14 juillet » à celle de l'année dernière. Les utilisateurs souhaitent identifier le pourcentage des ventes totales de la campagne à ce jour pour chaque jour de la campagne et comparer les régions et les produits les plus/les moins performants au quotidien.

## CONCLUSION

Ces nouvelles plates-formes massivement évolutives peuvent véritablement changer la donne dans le domaine de l'analyse prédictive. Quels avantages présentent-elles par rapport aux plates-formes d'entrepôt décisionnel et d'analyse prédictive actuelles ?

- Agilité permettant de provisionner et de réallouer des ressources informatiques massives à la demande, en fonction des priorités de l'entreprise.
- Possibilité d'analyser des Datasets à faible latence plus disparates et granulaires (Big Data) tout en conservant les nuances et relations détaillées dans les données pour fournir les informations détaillées différenciées qui permettent d'optimiser les performances de l'entreprise.
- Collaboration entre les services autour de projets métiers essentiels et dissémination rapide des bonnes pratiques et des résultats organisationnels.

## ÉTUDE DE CAS : HAVAS DIGITAL

### LA PUISSANCE DE L'ANALYSE PRÉDICTIONNELLE INTÉGRÉE, MOTEUR DE LA MODÉLISATION D'ATTRIBUTION COMPLEXE

Havas Digital et la division Data Computing Products d'EMC travaillent en collaboration en vue d'améliorer l'Artemis Analytics Lab, un projet de recherche et de développement couvrant le traitement analytique du Big Data à grande échelle dans le cloud privé. Ce projet permet d'accroître considérablement la compréhension par Havas Digital du comportement des utilisateurs et sa capacité à optimiser les campagnes marketing en conséquence.

Artemis Analytics Lab est un projet qui combine l'expertise d'Artemis en matière de marketing orienté données et l'environnement de base de traitement de données d'EMC® Greenplum pour fournir des fonctions optimales d'analyse numérique des données marketing et d'analyse prédictive intégrée dans la base de données.

« Pour la majorité des spécialistes du marketing, les informations démographiques sur les utilisateurs ne suffisent plus », explique Katrin Ribant, vice-présidente d'Artemis. « Nos clients veulent comprendre les aspects du comportement des utilisateurs qui sont uniquement identifiables dans le temps et déterminer ce qui se cache derrière de simples listes de transactions. Avec l'analyse prédictive avancée, EMC Greenplum® offre des fonctions d'analyse intégrées directement dans la base de données et qui s'appliquent à de très grands Datasets. Ainsi, nous pouvons générer de nouvelles informations détaillées enrichies sur le comportement des utilisateurs et aider les spécialistes du marketing à anticiper la réponse des utilisateurs aux nouvelles campagnes. »

L'un des principaux facteurs de différenciation d'Artemis est son mécanisme d'attribution sophistiqué qui détermine avec une grande précision l'impact relatif de la publicité sur les événements d'achat. En transférant la modélisation et les autres calculs vers la base de données et en utilisant EMC Greenplum, Havas Digital peut désormais proposer à ses clients un framework d'analyse prédictive d'attribution algorithmique quasiment en temps réel.

- Réduction des coûts pour tirer le meilleur parti des composants de traitement de base et analyser les masses de données (Big Data) en vue de traiter et d'exploiter les opportunités métiers qui ne pouvaient pas l'être d'une manière économique (voir pas du tout).

La plate-forme idéale pour l'analyse prédictive dispose des fonctions suivantes : une puissance de traitement massivement évolutive, la capacité d'exploiter des Datasets à granularité fine, un accès aux données à faible temps de latence et une intégration étroite entre l'entrepôt décisionnel et l'analyse prédictive. Correctement comprise et déployée, elle peut servir à résoudre des problèmes métiers complexes qui ne pouvaient pas l'être auparavant et à fournir à l'entreprise des analyses matérielles et décisionnelles.

## À PROPOS DE L'AUTEUR

Bill Schmarzo, Global Competency Lead chez EMC Consulting, compte plus d'une vingtaine d'années d'expérience dans le domaine des applications d'entrepôt décisionnel, d'aide à la décision et d'analyse prédictive. Il est l'auteur de la méthodologie « Business Benefits Analysis » qui lie les initiatives métiers stratégiques de l'entreprise à leurs exigences sous-jacentes en données et analyses. Il a également écrit plusieurs articles sur les applications de l'analyse prédictive en collaboration avec Ralph Kimball. Bill Schmarzo a occupé le poste de responsable du programme de formation aux applications d'analyse prédictive au Data Warehouse Institute.

EMC<sup>2</sup>, EMC, Greenplum et le logo EMC sont des marques déposées ou des marques commerciales d'EMC Corporation aux États-Unis et dans d'autres pays. Toutes les autres marques citées dans le présent document sont la propriété de leurs détenteurs respectifs.  
© Copyright 2011 EMC Corporation. Tous droits réservés. 11/05 EMC Perspective H8668.1