



TRANSFORMER L'ANALYSE  
DÉCISIONNELLE  
À TOUS LES NIVEAUX DE  
L'ENTREPRISE



# Le décisionnel pour tous

PAR SYBASE, AN SAP® COMPANY

# Le décisionnel pour tous

TRANSFORMER L'ANALYSE DÉCISIONNELLE  
À TOUS LES NIVEAUX DE L'ENTREPRISE

**Le décisionnel pour tous**  
Transformer l'analyse décisionnelle  
à tous les niveaux de l'entreprise

Publication par Sybase, an SAP Company  
One Sybase Drive, Dublin, CA 94568-7902, U.S.A.

Vous pouvez commander des exemplaires de ce Guide  
sur le site [sybase.com/analyticsguide](http://sybase.com/analyticsguide)

Copyright © 2012 Sybase, an SAP Company. Tous droits réservés.

Sybase et le logo Sybase sont des marques déposées de Sybase, Inc. ou de ses filiales. Le symbole ® désigne une marque déposée aux États-Unis d'Amérique. SAP et le logo SAP sont des marques ou des marques déposées de SAP AG en Allemagne et dans plusieurs autres pays. Toutes les autres marques appartiennent à leurs propriétaires respectifs.

Design : Boing Design



**Irfan Khan,**  
Vice-président exécutif et Directeur des technologies  
Sybase, an SAP Company

**À travers des faits et des exemples**, ce Guide démontre à quel point est exagérée l'anxiété largement exprimée dans le secteur du décisionnel à propos de l'arrivée de jeux de données très volumineux, de l'accueil d'un nombre beaucoup plus important d'utilisateurs et de l'ajout de types de données de plus en plus variées. Les craintes selon lesquelles les big data pourraient saper les performances et l'évolutivité des entrepôts de données décisionnels sont totalement infondées. Cela ne signifie pas que construire un entrepôt de données décisionnel à la fois réactif, complet et évolutif ne se fait pas sans relever quelques défis. Ces défis existent, mais ils ont été surmontés dans chacun des cas présentés ici.

Preuve à l'appui, ce Guide montre comment des entreprises de secteurs différents ont réussi à généraliser le décisionnel au plus profond de leurs activités. Il présente les résultats de recherches sur la valeur ajoutée du décisionnel, mais, plus important, il s'attarde sur plusieurs études de cas montrant à quel point le retour sur investissement du décisionnel est significatif pour les entreprises, en particulier celles qui voient augmenter leurs jeux de données et la population de leurs utilisateurs et qui doivent gérer différents types de données, des charges de travail plus lourdes et un nombre croissant de requêtes ad hoc.



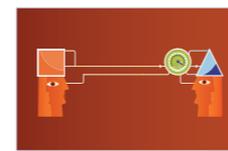
## 8 **Partie I - Big Data : un mensonge de taille**

- 11 Informatique : chronique d'une fin du monde annoncée
- 14 Quatre étapes pour assurer la qualité des données
- 16 Besoin accru de données
- 21 En matière d'analyse, il n'y a pas d'excuse qui tienne



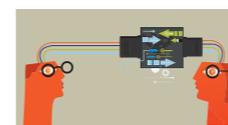
## 22 **Partie II - Le décisionnel pour tous**

- 26 Entrepôts de données : le talon d'Achille de l'entreprise ?
- 29 Une architecture dédiée à l'analyse décisionnelle



## 34 **Partie III - L'analyse décisionnelle sur le terrain**

- 37 Recevoir le message
- 39 Comment transformer les volumes de données gigantesques en atout de taille
- 42 Moins d'espace, plus d'informations pertinentes
- 43 La couverture du Cloud
- 43 Appliances : une envergure limitée
- 44 Utilisateurs nomades
- 45 Plus d'utilisateurs, moins de maux de tête
- 47 Une simple portion fait office d'échantillon



## 48 **Partie IV - Meilleures pratiques et retour sur investissement de l'analyse décisionnelle**

- 51 Le potentiel de la démonstration de faisabilité
- 53 Les sept préceptes d'une analyse décisionnelle réussie
- 54 Au-delà de la démonstration de faisabilité
- 56 Calculer le retour sur investissement de l'analyse décisionnelle
- 58 De la recherche à la réalité
- 59 La passion des performances et des bénéfices

60 Sources

62 Ressources



Big data :  
un mensonge de taille

## Le ciel est en train de nous tomber sur la tête, une nouvelle fois.

Cette fois-ci, il pleut des big data sur la tête des responsables informatiques. Il faut avoir peur, très peur.

Comme dans ce film où Chicken Little rencontre des extra-terrestres, la crise des volumes gigantesques de données est une fiction, un mensonge de taille.

Mythe urbain lancinant, le sujet des big data est omniprésent dans les médias ces derniers temps. Il est impossible d'y échapper. Une recherche sur « big data » dans Google donne des millions de résultats. Des articles importants ont été consacrés au sujet non seulement dans la presse informatique, mais aussi dans la presse généraliste (notamment dans le New York Times et le Christian Science Monitor, aux États-Unis). On trouve même une entrée « Big data » dans Wikipedia. Et ce n'est pas tout : des conférences réputées dédient de nombreuses sessions à ce sujet et le site GigaOm a même lancé un événement annuel exclusivement sur ce thème — le premier est d'ailleurs déjà accessible en ligne<sup>1</sup>. Cette déferlante d'informations a laissé penser à beaucoup que le secteur de l'informatique allait être submergé. Selon un cabinet d'analyse, « La croissance des données est le principal défi posé aux entreprises majeures en matière d'infrastructure matérielle de leur data center »<sup>2</sup>.

Avec un tel battage, il n'est pas étonnant de voir les hauts responsables informatiques scruter nerveusement leur infrastructure décisionnelle et s'inquiéter de savoir si elle résistera sous l'assaut violent d'effrayants jeux de données volumineux. Ils doivent entendre ce conseil : détendez-vous. Le véritable problème ne réside pas dans les big data.

Cela ne signifie pas que le volume d'informations qui pénètre dans l'entreprise ne grandisse pas, bien au contraire. D'ici à 2020, les entreprises auront à gérer un volume de données multiplié par 35, selon une étude de marché conduite par IDC<sup>3</sup>. En fait, les big data ont toujours été une gageure, mais il leur manquait jusqu'alors une appellation suffisamment attrayante pour que les journalistes, analystes et autres experts puissent s'en gargariser. ■■■



Informatique :  
chronique d'une  
fin du monde  
annoncée

**D**ÈS LES ANNÉES 50 le problème du volume des données s'est posé dans les entreprises. IBM a compris qu'à défaut d'une avancée technologique majeure, le secteur des mainframes qu'elle commençait à développer allait droit dans le mur : le volume des données transactionnelles de ses systèmes grandissait plus vite que la capacité en mémoire autorisée par les technologies d'alors.

En 1956, IBM a donc inventé la première unité de disque dur magnétique permettant de stocker des informations à l'extérieur de la mémoire centrale de l'ordinateur<sup>4</sup>. À environ 10 000 dollars US le mégaoctet (soit 7300 euros), l'imposante machine mise au point par IBM, d'une capacité de 4,4 Mo, était constituée de 50 disques d'un diamètre de 60 cm chacun, rangés dans une armoire de la taille d'un réfrigérateur. On pensait qu'elle serait capable de résoudre les problèmes d'évolutivité des données de l'époque. Ce ne fut pas le cas.

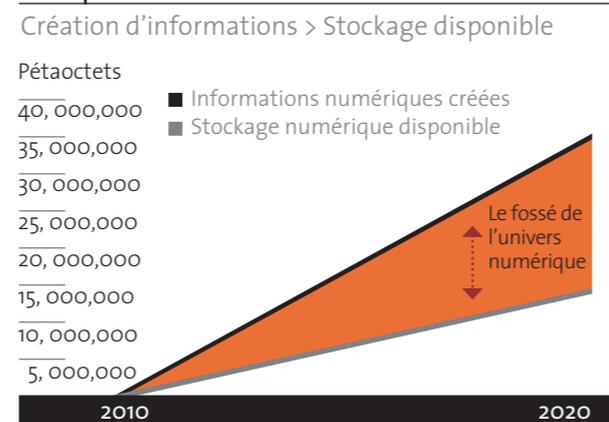
À mesure que la capacité grandissait, les données se sont démultipliées. Il a donc fallu plus d'ordinateurs pour traiter toutes les informations. Leur nombre a donc bondi de 600 en 1956 à 30 000 en 1966, puis à 400 000 en 1976, à travers le monde. Et avec l'arrivée des micro-ordinateurs en 1984, il a explosé pour atteindre les 6 millions d'unités<sup>5</sup>. On estime aujourd'hui à 3 milliards le parc mondial d'ordinateurs. Rien qu'en 2010, la production a été de 352 millions d'unités<sup>6</sup>. Quant au volume de données généré par ces appareils, il est étourdissant : en 2009, IDC a estimé à 800 000 pétaoctets la quantité de données générées et stockées dans le monde, soit un total de 35 zettaoctets en moins de 10 ans.

À chaque étape du parcours, les observateurs ont déploré que les données créées par toutes ces machines finissent par dépasser les individus et les entreprises qui les employaient. En 1970, le best-seller d'Alvin Toffler, « Le choc du futur », a popularisé le

concept de « surcharge d'information » pour expliquer pourquoi les personnes et les entités peinaient à prendre des décisions face à des volumes trop importants de données. En 1986, le détracteur des technologies Theodore Roszak écrivait : « Un excès d'informations pourrait en réalité envahir les idées, laissant l'esprit... distrait par des faits stériles et sans rapport les uns avec les autres, perdu au milieu d'amas de données informes »<sup>7</sup>. Et les craintes relatives aux flux d'informations n'étaient pas seulement l'apanage d'écrivains de renom, elles étaient également relayées par les fondus d'informatique. Par exemple, en 1990, à l'occasion d'une conférence de l'IEEE sur la capacité des disques et les limitations des performances, l'organisateur d'une session avait été invité à intituler sa thématique « La crise dans le stockage de masse ».

On nous dit que la crise actuelle est différente. Selon un rapport, « l'échelle et la portée des changements entraînés par les big data constituent un point d'inflexion à partir duquel leur expansion va bondir, soutenue par l'accélération et la convergence d'une série de technologies »<sup>8</sup>. Une autre étude indique que le volume des données créées pourrait dépasser la capacité de stockage avant 2020<sup>9</sup>.

**Fig. 1** | Le fossé se creuse



Source: Étude d'IDC, « Digital Universe Study », commanditée par EMC, mai 2010

En dépit de ces annonces de catastrophe imminente due aux déferlements de données, le secteur de l'informatique a régulièrement tenu bon. Il a toujours lutté contre la fin du monde des informations sans cesse annoncée en perfectionnant notre infrastructure informatique, la rendant plus rapide, meilleur marché et moins encombrante, tout en augmentant sa capacité. Prenons simplement l'exemple d'un élément de cette infrastructure - l'unité de disque dur - pour illustrer la stratégie mise en œuvre pour éloigner le spectre des big data pendant toutes ces années :



**La capacité des unités de disque dur est passée** de 3,75 Mo à plus de 1 To, soit un gain de 270 000 fois la capacité initiale.



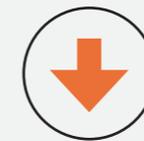
**La taille des unités de disque a fondu**

de presque 2,5 m<sup>3</sup> au départ (encombrement d'un réfrigérateur américain) à 56 cm<sup>3</sup> (version compacte des disques durs de 2,5", c'est-à-dire la taille d'un paquet de cartes), soit une réduction de plus de 44 000 fois par rapport à la taille initiale.



**Les prix ont chuté**

de la somme astronomique de 15 000 dollars US (env. 11 000 euros) à moins de 0,0001 dollar US par mégaoctet — c'est-à-dire 100 dollars US ou 74 euros par téraoctet —, soit 150 000 000 fois moins !



**Le temps d'accès moyen a été réduit**

de 40 fois, s'accroissant d'un peu plus de 0,1 seconde à quelques millièmes de seconde<sup>10</sup>.

Au-delà de ces avancées incroyables concernant les modestes unités de disque, les fabricants informatiques ont amélioré l'ensemble des composants matériels de l'assemblage de calcul, des performances des microprocesseurs à la vitesse des bus internes. Les crises répétées autour des big data ne pouvaient cependant pas se résoudre uniquement par l'amélioration du matériel, surtout dans le domaine du décisionnel. ■■■

# Quatre étapes pour assurer la qualité des données

POUR DÉPLOYER UN ENTREPÔT DE DONNÉES DÉDIÉ À L'ANALYSE DÉCISIONNELLE, IL EST IMPÉRATIF DE S'ASSURER DE LA QUALITÉ DES DONNÉES. IL EXISTE DES TECHNIQUES ÉPROUVÉES POUR CELA : DU NETTOYAGE DES DONNÉES AU RECOURS À UNE GESTION DES DONNÉES DE RÉFÉRENCE. QUOI QU'IL EN SOIT, LA DÉMARCHE DOIT NÉCESSAIREMENT INCLURE UNE VÉRIFICATION DES ÉLÉMENTS SUIVANTS EN INTERNE :

1.

**Latence des données** : tenez compte des trois dimensions de la latence des *données* dans votre entreprise : la périodicité des données, la durée de l'événement et le délai de la prise de *décision*.

2.

**Corrélation des données** : collaborez avec les utilisateurs de l'entreprise pour déterminer le contexte des données et établir les relations entre les différents jeux de données utilisés, tout en conservant à l'esprit les taux de croissance des données et les sources de duplication.

3.

**Libre-service** : déterminez comment les utilisateurs autorisés pourront contrôler les données qu'ils utilisent pour leurs requêtes, sans impact sur le service informatique ni sur d'autres ressources.

4.

**Responsable des données** : confiez à un cadre supérieur la responsabilité de gouverner les données en assurant leur qualité opérationnelle, dans le respect de la démarche de gouvernance choisie par l'entreprise (son titre est facultatif).

La valeur accordée à la qualité des données ne sera jamais exagérée. Une étude de l'université du Texas (University of Texas) conclut que des données de qualité ou efficaces se répercutent à la fois sur les ventes et sur les bénéfices des entreprises. Leurs différents effets sont :

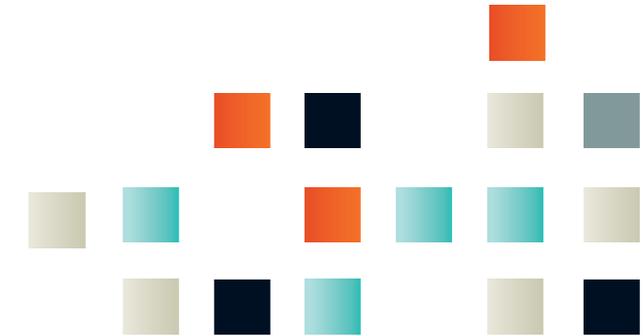
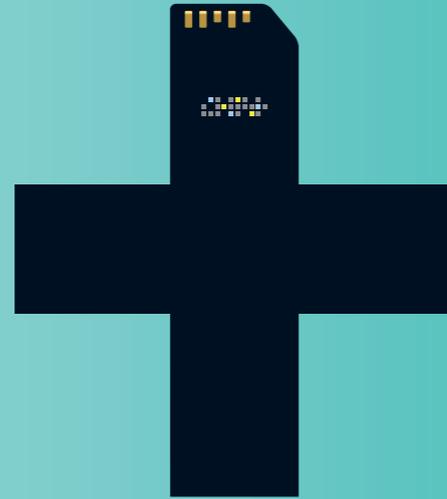


**Impacts financiers** dans des domaines comme la productivité des salariés, le rendement des capitaux propres, le rendement du capital investi et le rendement des actifs.

**Impacts sur la clientèle** dans des domaines comme la capacité à innover pour dégager du chiffre d'affaires sur les nouveaux produits et pour développer la base de clients existante.

**Impacts opérationnels** dans des domaines comme l'utilisation des actifs, la précision de la planification et des prévisions et le respect des délais de livraison des produits ou de fourniture des services<sup>11</sup>.

## Besoin accru de données



**L**A PREMIÈRE CONFÉRENCE internationale sur la découverte des connaissances et l'exploration des données (International Conference on Knowledge Discovery and Data Mining) s'est tenue en 1995 à Montréal. Ce symposium trouvait sa justification dans l'opinion dominante selon laquelle d'importantes quantités de données susceptibles d'améliorer la vie en société et dans l'entreprise étaient ignorées tout simplement parce que les jeux de données étaient trop imposants pour être analysés.

Dans la brochure de la manifestation, les organisateurs avaient écrit : « Ayant atteint des tailles défiant un examen, même partiel, par l'œil humain, les bases de données et les collections de jeux de données modernes débordent littéralement les utilisateurs. Cette déferlante de données est un phénomène qui se reproduit dans des secteurs aussi différents que l'analyse des données scientifiques, la médecine et la santé, l'entreprise et le marketing, et les marchés financiers »<sup>12</sup>. L'un des objectifs essentiels de la Conférence était de partager les méthodologies, modèles statistiques, techniques d'apprentissage machine et autres approches non matérielles - alors émergentes -, pour résoudre le problème de l'exploration de volumes de données toujours plus importants.

Les débats ont notamment porté sur l'utilisation de jeux de données volumineux pour repérer les fraudes et pour découvrir des traitements prometteurs contre le VIH. Dans les deux cas, l'utilisation conjointe de jeux de données volumineux et de techniques d'analyse avancées ont permis de réaliser de grands progrès.

C'est aussi en 1995 que la base de données en colonnes fait son apparition. Radicalement différente de la base de données relationnelle classique, conçue dans les années 60 pour répondre à tous les besoins, l'approche en colonnes a été un grand pas vers la création d'un entrepôt de données pour le décisionnel (cf. la deuxième partie).

Les entreprises qui utilisent aujourd'hui une infrastructure d'analyse en colonnes n'ont aucun souci à se faire concernant les big data. Leurs activités ont au contraire tout à y gagner. Les jeux de données massifs, y compris ceux contenant des types de données variables comme les données non structurées, se prêtent à l'analyse dans un entrepôt de données en colonnes. Non seulement ils s'y prêtent, mais leurs performances sont supérieures en termes de vitesse et leur évolutivité leur permet d'inclure d'emblée autant d'utilisateurs et de données que nécessaire pour mener à bien la mission.

L'objectif ne réside pas dans les capacités ou ►►►

## Les quatre facteurs

de l'analyse décisionnelle en entreprise

### VOLUMES DE DONNÉES TOUJOURS PLUS IMPORTANTS

-- À terme, les données dépasseront les capacités de stockage

### EXPLOSION DES POPULATIONS D'UTILISATEURS

-- Des millions d'explorateurs de données potentiels

### DONNÉES NON STRUCTURÉES À DÉFRICHER

-- La multiplicité des données devient la norme

### ACCÉLÉRATION

-- La pénétration des données prend de la vitesse

les gains de performances envisagés, ni dans la possibilité de déborder des big data en portions plus petites pour les répartir dans des datamarts. Il ne s'agit pas non plus de restreindre les populations d'utilisateurs, ni d'exercer une plus grande maîtrise des charges de travail ou d'éliminer les requêtes ad hoc ou, encore moins, de limiter les fonctions d'analyse face aux big data.

Non, l'objectif consiste à dénicher les big data à l'intérieur comme à l'extérieur de l'entreprise et à en tirer parti pour faire pénétrer le décisionnel au plus profond de celle-ci et au bout du compte optimiser sa réactivité, sa compétitivité et sa rentabilité.

Illustrons ce propos par l'exemple de comScore. Ce prestataire de cloud computing, expert des solutions et services d'analyse pour le marché de l'e-commerce, a compris dès le début de son activité que le marketing sur Internet était en train de changer d'orientation pour privilégier la rentabilité sur le nombre de visiteurs. Sa plateforme Customer Knowledge Platform offre une vision à 360° du comportement et des préférences des clients qui naviguent d'un site à l'autre. Son service surveille les habitudes de navigation et d'achat sur chacun des sites visités par les consommateurs qui ont accepté que leur comportement soit analysé.

Comme des millions d'internautes ont choisi de participer, les volumes de données collectés sont énormes. De fait, comScore fait tourner ses fonctions d'analyse sur plus de 40 téraoctets de données compressées, auxquels viennent s'ajouter chaque semaine 150 gigaoctets.

Malgré ce volume de données, les temps de réponse des requêtes sont exceptionnels. Comme le confirme Ric Elert, Vice-président chargé de

### Profil de structure en colonnes | comScore

#### 1999

Fondée en 1999

#### 3 millions

Analyse plus de 3 millions de sites Web

#### 2 millions

Le panel de comScore inclut plus de 2 millions de consommateurs dans 170 pays

#### Cloud computing

Prestataire de cloud computing spécialiste du comportement des consommateurs en ligne

#### 150 gigaoctets

Jeux de données volumineux augmentant de 150 gigaoctets par semaine

#### 40 téraoctets

40 téraoctets de données compressées - taux de compression de 40 %

l'ingénierie chez comScore, l'entreprise peut ainsi « explorer les données et produire des résultats pour nos clients bien plus rapidement. Cela permet à ces derniers de commercialiser plus efficacement leurs produits et d'augmenter leur chiffre d'affaires ».

En outre, la technologie de stockage en colonnes permet à comScore d'atteindre des taux de compression de 40 %. L'entreprise indique d'ailleurs qu'une approche classique aurait fait grimper ses coûts de stockage.

« La compression est essentielle à nos yeux car nous sommes soumis à des afflux massifs de données », souligne Scott Smith, Vice-président chargé des entrepôts de données. « La quantité de données dont nous disposons est affolante. Elle dépasse ce que la plupart des gens seront jamais amenés à voir ».

À l'instar de comScore, Suntel se développe rapidement. Avec ses 500 000 abonnés, il est même l'opérateur de télécommunications qui

enregistre la plus forte croissance au Sri Lanka. En s'appuyant sur les toutes dernières technologies, une approche novatrice et un engagement sans précédent en matière de service, cette entreprise a personnalisé des solutions de télécommunications pour satisfaire les exigences les plus complexes de ses clients. Forte de son inventivité alliée au déploiement d'une technologie d'avant-garde, Suntel ambitionne de devenir un acteur majeur de la transformation de la société sri-lankaise en une société d'information fondée sur le savoir.

Dans un tel contexte, la base de données relationnelle dont disposait Suntel n'offrait pas l'excellence nécessaire pour lui permettre d'atteindre ces objectifs. En particulier, les requêtes s'exécutaient à une lenteur inacceptable.

« Nous en étions arrivés au point », explique Tariq Marikar, Directeur des technologies de l'information et de la fourniture des solutions,

►► « où notre base de données de production accusait une surcharge de 20 % tout à fait inacceptable pour nous. En plus, nous souhaitions pouvoir générer des rapports et exécuter des requêtes sur des années de données historiques et pas seulement sur quelques mois. Nous savions que la solution passait par la création d'un référentiel distinct - un entrepôt de données - spécifiquement conçu et dédié au reporting et aux analyses ».

En optant pour un entrepôt de données en colonnes conçu pour effectuer des analyses avancées, Suntel a atteint son but. « Il est plus qu'essentiel pour notre activité d'avoir une visibilité sur les grands volumes de données historiques », précise Tariq Marikar. À l'instar de comScore, avec l'approche en colonnes, « la capacité de compression nous a permis de réduire d'environ un tiers l'espace occupé par les données résidant dans notre base de production ».

Il ajoute à cela qu'avec ce système, « l'évolutivité

signifie que nous pourrions tirer parti de la valeur de ces données à l'avenir. Par exemple, nous étudions actuellement comment exploiter cette richesse de données en vue de développer des moyens pour personnaliser notre interaction avec des clients d'envergures différentes et mettre en œuvre des programmes de vente croisée et poussée de nos services ».

En Espagne, l'entrepôt de données en colonnes d'Airtel Vodafone organise les informations en fonction de l'écosystème de l'entreprise. De cette façon, même si de nombreux services différents utilisent les mêmes données, la cohérence et l'intégrité des informations sont préservées. L'entrepôt de données convertit les données en savoir, transformant par le biais d'une interface les faits du monde réel en précieuses données décisionnelles. Cette capacité à analyser et prévoir avec précision l'activité de ses clients est vitale dans la stratégie globale d'Airtel Vodafone.

La charge des demandes de requête est intense sur l'entrepôt de données en colonnes, mais le système la gère facilement. Plus de 1000 salariés utilisent l'entrepôt de données pour mener des analyses multidimensionnelles, plus exigeantes en termes d'utilisation du système. Lorsque ces analyses s'effectuent sur des structures d'informations conçues spécifiquement à cet usage, cela permet d'extraire les données relatives aux clients, infrastructures et processus de l'entreprise. Les utilisateurs peuvent ainsi créer des processus de modélisation et de simulation incluant des techniques qui complètent l'analyse.

Les techniques d'exploration des données (data mining) permettent également d'extraire les informations sur les schémas comportementaux des clients. Chez Airtel Vodafone, le personnel qui est en contact avec la clientèle peut entrer quotidiennement les informations collectées pour qu'elles soient intégrées avec celles déjà stockées dans l'entrepôt. Ces données sont ensuite combinées et converties en structures d'informations pour les demandes de renseignements.

L'entrepôt de données en colonnes permet d'accéder aux informations en fonction des besoins du workflow plutôt que sur la base de la structure hiérarchique de l'entreprise. Cela favorise la productivité et l'efficacité des salariés. Ainsi, les utilisateurs du service marketing accèdent aux mêmes informations que ceux du service financier, bien que depuis des angles différents et à l'aide d'analyses différentes. L'environnement de l'entrepôt de données comprend les bases de données du marketing, les systèmes d'assistance téléphonique, le service client, les données statistiques relatives au réseau GSM, les systèmes de facturation, les collections et les extractions, ainsi que toutes les informations logistiques. ■■■



## En matière d'analyse, il n'y a pas d'excuse qui tienne

**CES EXEMPLES** ne constituent qu'un petit échantillon d'entreprises qui savent par expérience qu'elles n'ont rien à craindre des big data et qu'au contraire, elles doivent les exploiter. Dans des sections ultérieures de ce Guide, vous trouverez d'autres exemples similaires d'entreprises qui ont réussi. Ils seront l'occasion de mettre l'accent sur d'autres aspects de l'actualité et de la technologie des fonctions d'analyse avancée pour l'entreprise.

Aujourd'hui, le secteur de l'analyse décisionnelle ne peut avancer aucune excuse quant aux volumes gigantesques de données. Idem concernant la mise à l'échelle des entrepôts de données analytiques pour prendre en charge des milliers d'utilisateurs. Pas d'excuse non plus concernant l'exécution de fonctions d'analyse sur des types de données variables provenant de toutes les sources de données imaginables, comme la manne d'informations non structurées provenant des sites des réseaux sociaux, par exemple.

Voyons les choses en face : le secteur de l'analyse décisionnelle ne peut plus agiter le spectre des big data car nous savons à présent qu'il ne s'agit que d'un mythe, d'un mensonge de taille. ■■■

### Profil de structure en colonnes | Airtel Vodafone

#### 2000

Fondée en 2000 à la suite de la fusion des actions Airtel Móviles détenues par British Telecom et Vodafone

#### 14 millions

14 millions de clients  
Deuxième fournisseur espagnol de services de téléphonie mobile

#### 1995

Airtel Móviles avait été fondée en 1995

#### 1000+

Plus de 1000 utilisateurs de l'entrepôt de données

#### 100%

Intégration totale des données entre tous les services

#### Multidimensionnelle

Analyse multidimensionnelle  
- Clients  
- Infrastructure  
- Processus métiers

# Le décisionnel pour tous



**C**OMME NOUS L'AVONS VU, comme nous l'avons vu dans la partie précédente, les entreprises dont l'activité repose sur des fonctions analytiques partagent la même problématique : leurs responsables informatiques considèrent que l'environnement décisionnel doit gérer des quantités croissantes d'informations et optimiser les temps de réponse des requêtes, tandis que les populations d'utilisateurs grandissent parallèlement. Elles savent que la démultiplication des données va attirer un nombre plus important d'utilisateurs qui auront de bonnes raisons de les interroger dans le cadre de leur mission. De fait, le problème du « nombre important d'utilisateurs » peut se révéler plus insurmontable pour certains fournisseurs d'entrepôts de données (et pour leurs clients) que celui des big data.

Dans une certaine mesure, n'importe quel système peut être optimisé pour mieux prendre en charge les jeux de données, y compris les

plus volumineux. En revanche, les utilisateurs sont moins faciles à gérer. Il est en effet impossible de prévoir les requêtes qu'ils exécutent, la quantité d'activité qu'ils génèrent. Ce sont les « jokers » de l'environnement analytique. Cependant, si le système n'est pas dimensionné pour répondre à leurs exigences, vous ne serez pas en mesure de mettre le décisionnel à la disposition de tous ceux qui en ont besoin dans l'entreprise.

Pour des entreprises comme Airtel Vodafone, comScore et SunTel, le décisionnel à la portée de tous n'est pas une chimère. Ces entreprises sont capables de mettre le décisionnel à la disposition de populations d'utilisateurs en rapide expansion parce qu'en choisissant un décisionnel adapté à leur activité, elles ont pris le recul nécessaire pour remettre en question l'approche basée sur une solution universelle.

Dans de très nombreux cas, une technologie de

base de données clé en main est parfaitement adaptée. Ainsi, un entrepôt de données d'entreprise générique sera d'une aide précieuse si les utilisateurs sont en nombre restreint et souhaitent disposer de rapports réguliers et d'informations sur les perspectives stratégiques, sans aucune contrainte de délai ou de performance.

Toutefois, il serait vain de tenter de bricoler une solution à partir d'un entrepôt de données d'entreprise générique si l'objectif est d'offrir des outils analytiques à une importante population d'utilisateurs grandissant sans cesse, avec des jeux de données dont la taille et la diversité ne font qu'augmenter et des requêtes ad hoc toujours plus nombreuses qui pèsent sur le système. En effet, les exigences d'un environnement analytique sont telles qu'il est inenvisageable de compenser les limites inhérentes à un entrepôt de données d'entreprise en ajoutant simplement un module ou

en consacrant un mois ou deux à l'optimisation du système. Cette stratégie mène droit dans l'impasse.

Pourquoi cette réponse aux besoins d'analyse à l'échelle de l'entreprise est-elle insuffisante? N'était-ce pas précisément l'objectif initial de l'entrepôt de données d'entreprise? Malgré des milliards d'euros investis dans les entrepôts de données d'entreprise, pourquoi le Gartner Group prédit-il pour l'année 2012 que plus de 35 % des 5000 premières entreprises mondiales échoueront à prendre des décisions éclairées menant à des changements efficaces pour leur activité et leur marché?<sup>13</sup>

Pour répondre à cette question, il faut s'intéresser de nouveau aux aspects technologiques et s'interroger sur les raisons pour lesquelles l'architecture des logiciels de votre entrepôt de données est si essentielle. C'est en effet sur elle que repose la promesse du décisionnel à la disposition de tous. ■■■

## L'analyse décisionnelle fait travailler différemment



**Cadre dirigeant**  
Visibilité instantanée  
des sujets petitset  
grands



**Responsable**  
Gestion plus avisée  
des programmes



**Analyste**  
Plus de données  
pour approfondir  
plus rapidement ses  
connaissances



**Expert en  
marketing**  
Campagnes  
basées sur les  
mesures



**Commercial**  
Ciblage plus rapide  
des clients à forte  
marge



**Opérationnel**  
Contrôles plus  
stricts, production  
mieux maîtrisée



**Technicien sur  
le terrain**  
Problèmes évalués,  
donc résolus, plus  
rapidement



**Partenaire**  
Moins de  
surprises sur  
la chaîne  
logistique



# Entrepôts de données : le talon d'Achille de l'entreprise ?



**C**OMME LE FAIT REMARQUER FRANK HAYES, observateur de longue date du marché, dans un article intitulé « The Story So Far »<sup>14</sup>, le concept d'entrepôt de données remonte à plus de 25 ans, bien avant qu'un nom lui ait été attribué. Il semblait alors concevable de capturer toutes les données pertinentes pour une activité dans un système de base de données transactionnel, générique et organisé en lignes. Le concept de « big data » était totalement inconnu et les utilisateurs susceptibles d'exécuter des fonctions d'analyse sur le système constituaient une population très limitée. Très logiquement, les services informatiques ont donc investi lourdement dans des entrepôts de données d'entreprise pour constituer leur principale infrastructure décisionnelle.

Ce comportement a engendré un problème fondamental, comme le souligne l'expert en décisionnel Colin White. « Initialement utilisés pour résoudre les problèmes de conception des systèmes transactionnels », les entrepôts de données d'entreprise ont été généralisés par les fournisseurs qui en ont étendu la portée pour « faire rimer les concepts du décisionnel et des entrepôts de données ». À cet égard, Colin White conclut : « C'est une croyance fautive dont il faut s'éloigner »<sup>15</sup>. Examinons à présent l'une des faiblesses de cette approche qui consiste à

répondre aux exigences du décisionnel analytique par une technologie d'entrepôt de données d'entreprise générique.

Votre entrepôt de données d'entreprise universel repose sur une base de données organisée en lignes (cf. figure 2 sur la page suivante « Trois types de stockage des données analytiques »). Cette configuration convient à merveille au traitement basé sur les transactions puisque l'objectif principal de l'application est d'enregistrer un événement. Prenons l'exemple d'un client X qui achète un widget. Les attributs de la transaction sont inclus dans leur intégralité au moment où la base de données écrit l'événement dans le système de stockage : données sur le client, informations tarifaires, spécifications relatives au widget, heure et autres détails, l'ensemble est capturé dans la transaction enregistrée. Inconvénient : à chaque fois qu'un analyste veut explorer un sous-ensemble de ces attributs, l'entrepôt de données d'entreprise universel doit lire tous les attributs en mémoire avant d'appliquer une quelconque fonction d'analyse. Dans ces conditions, si vous explorez la relation entre l'heure et le type de widget vendu, par exemple, et que vous souhaitez utiliser ces données conjointement avec des sources d'information externes, le système devra extraire de la transaction toutes les informations sur les clients et les tarifs en même temps que les données pertinentes. Cette ►►►

Fig. 2 | Trois types de stockage des données analytiques

**1 En colonnes**

Avec des données stockées en colonnes, il est possible d'obtenir un niveau de compression élevé, donc de réduire radicalement l'espace disque occupé. Cela accélère en outre le temps de réponse des requêtes car les données sont lues plus rapidement par les opérations de base de données, par comparaison avec une approche classique. Les données sont indexées automatiquement au moment de leur chargement dans la base.

**2 En mémoire**

Cette approche qui consiste à stocker toutes les données dans la mémoire vive de l'ordinateur est certes très réactive, mais inefficace en environnement 32 bits en raison de la capacité relativement restreinte de la mémoire. Et si elle est envisageable sur les systèmes 64 bits qui disposent d'une vaste mémoire vive, elle n'en reste pas moins très onéreuse.

**3 En lignes**

Lorsque les données sont stockées en lignes sur le disque, elles se prêtent à un enregistrement sur la base des transactions, mais leur indexation est difficile. Cette approche n'apporte pas le niveau de performance et d'évolutivité des architectures plus récentes.

►►► plongée dans des données tout à fait inutiles aura pour effet de ralentir les temps de réponse.

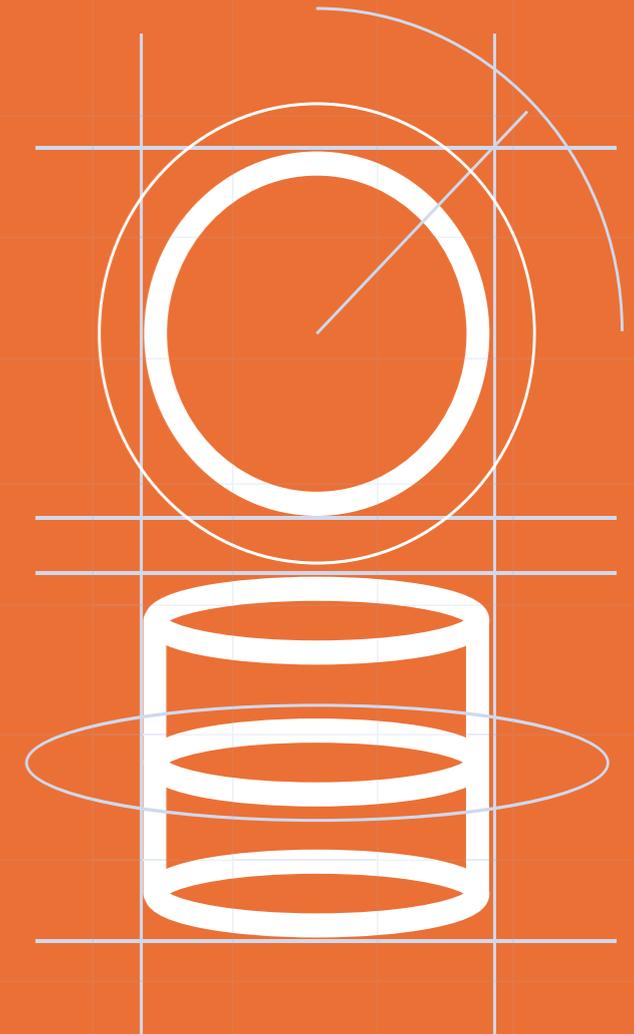
Évidemment, les partisans d'un entrepôt de données d'entreprise générique soutiendront qu'il existe des techniques pour pallier ces inconvénients. La méthode la plus répandue consiste à créer et stocker des index séparés des données pour compenser la lenteur des temps de réponse. Attention toutefois, car si certaines astuces d'optimisation permettent d'améliorer les performances des entrepôts de données d'entreprise universels, elles entraînent des coûts et temps passés non négligeables et complexifient la gestion du système. De plus, le manque de souplesse de ce genre de méthode limite les types de requête qu'il est possible de générer lorsque l'équipe informatique n'a pas créé les index appropriés. Même correctement configurées, ces solutions génériques n'offrent pas des résultats comparables à ceux des moteurs d'analyse

spécifiques en termes de performance et d'évolutivité. L'un des pionniers des bases de données relationnelles organisées en lignes a même décrété cette technologie dépassée<sup>16</sup>.

Les entrepôts de données ont été conçus pour les transactions. Comme ils renfermaient la majorité des informations de l'entreprise, les fonctionnalités d'analyse sont venues s'y greffer. Ils sont ainsi devenus des moteurs d'analyse par accident. Du point de vue de l'utilisateur, des accidents pourraient d'ailleurs bien se produire. En effet, dans une perspective de performance et d'évolutivité, dès lors que les temps de réponse deviennent essentiels pour mener à bien une mission ou que l'aptitude à prendre en charge des nouveaux utilisateurs par centaines, voire par milliers, constitue un avantage déterminant sur la concurrence, on peut dire que ces entrepôts de données qui faisaient l'affaire jusqu'à présent risquent de devenir le talon d'Achille de l'entreprise. ■■■

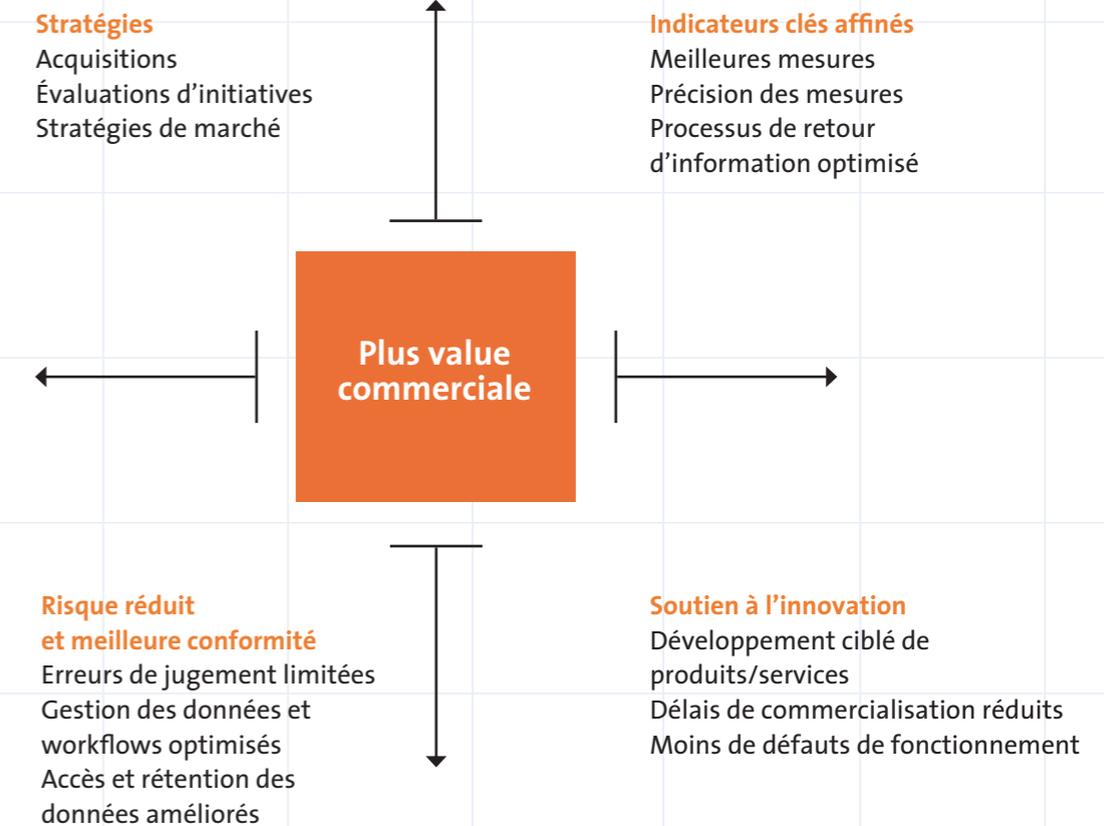
# Une Architecture dédiée à l'analyse décisionnelle

SI LA CONCEPTION DE LA BASE DE DONNÉES  
RELATIONNELLE CLASSIQUE ET FIABLE NE  
SATISFAIT PAS LES EXIGENCES DE L'ANALYSE  
DÉCISIONNELLE D'AUJOURD'HUI, VERS  
QUELLE SOLUTION SE TOURNER ? QUELLES  
SONT LES CARACTÉRISTIQUES ESSENTIELLES  
DE L'ARCHITECTURE D'UN ENTREPÔT DE  
DONNÉES DÉCISIONNEL ?



► **P**OUR COMMENCER, comme évoqué dans la première partie, vous devez rechercher un moteur d'analyse spécialisé pour la tâche. Un tel système stocke probablement les données en colonnes. L'un des principaux atouts de ce type d'architecture réside dans sa capacité à compresser les données. Cette approche n'a rien de théorique car des applications existent bel et bien dans le monde réel. Par exemple, la filiale allemande de Citigroup indique stocker quatre années de transactions financières, soit 13,2 millions d'opérations, dans une base de données en colonnes compressée à 43 %. En comparaison, l'approche classique en lignes ferait grossir chacune des opérations stockées de 4 %. De la même façon, lorsqu'elle analyse des activités commerciales en ligne, la société comScore obtient un taux de compression de ses collections de données de 40 % grâce à l'approche en colonnes. La compression peut même atteindre 70 % dans d'autres entreprises qui appliquent cette méthode.

## Avantages de l'analyse décisionnelle en entreprise



►►► La compression n'est qu'une des caractéristiques de l'organisation en colonnes. Si votre système offrait uniquement celle-ci, vous ne pourriez pas bénéficier de fonctions d'analyse avancées. C'est pourquoi l'architecture globale doit pouvoir gérer les données compressées en mode parallèle, sur l'ensemble d'un système appelé massivement parallèle (ou MPP pour Massively Parallel Processing), ce qui constitue un véritable atout du point de vue des performances. Mais si tous les systèmes de stockage en colonnes indexent automatiquement les données compressées, leur architecture peut différer. Certains ajoutent une métacouche par-dessus les données pour les faire fonctionner avec le système MPP sous-jacent. Cette métacouche ralentit les performances brutes, restreint la capacité de l'entreprise à gérer les requêtes ad hoc, limite la croissance de la population d'utilisateurs et peut même présenter des défaillances en présence de big data.

La structure en colonnes d'une architecture analytique spécialisée permet également d'accélérer les déploiements. Comparée à un entrepôt de données

d'entreprise générique dépassé - qui requiert plusieurs mois d'indexation manuelle pour optimiser les données -, la base de données en colonnes indexe automatiquement les données à mesure qu'elles sont chargées dans le système.

L'architecture spécialisée offre une évolutivité bien supérieure à celle des bases de données en ligne classiques. Elle permet donc une approche de type « shared-everything », où tout est partagé, par opposition à celle de type « shared-nothing ». Un entrepôt de données MPP qui utilise une approche « shared everything » peut gérer et équilibrer dynamiquement les charges de travail des requêtes sur l'ensemble des nœuds de la grille. Il se sert d'un rééquilibrer de charge de travail automatique agissant très efficacement pour éviter les conflits entre utilisateurs des ressources système. Cela lui permet d'assurer un niveau de performances élevé et prévisible face à un large spectre de charges de travail concurrentes.

Une architecture de type « shared-everything » peut faire évoluer le volume de stockage des données et le volume des demandes émanant d'utilisateurs

concurrents indépendamment l'un de l'autre, assurant ainsi un équilibre optimal des performances, en toute simplicité. Cela vous permet d'ajouter sans cesse des données et des utilisateurs sans alourdir les charges de travail ni entraver le fonctionnement pour autant.

Autre aspect de l'architecture analytique spécialisée : sa capacité à exécuter des fonctions d'analyse embarquées dans la base de données. En clair, toutes les données pertinentes dans le cadre d'une charge de travail doivent être disponibles pour traitement sans que le système doive les récupérer à l'extérieur.

Les allers-retours de données sont néfastes pour les performances. C'est pourquoi les temps de réponse sont largement meilleurs lorsque les fonctions d'analyse se trouvent à l'intérieur même de la base.

Une architecture analytique spécialisée est conçue dès l'origine pour s'accommoder de tous les types de données, ce qui n'est pas le cas des entrepôts de données d'entreprise génériques conçus à une époque où seules les données structurées se prêtaient aux applications décisionnelles. Aujourd'hui, n'importe quel moteur d'analyse sérieux doit pouvoir exécuter sans effort des fonctions décisionnelles sur de grands volumes de données non structurées provenant de toutes sortes de sources (cf. l'encart « Données non structurées : une véritable mine d'or »). IDC estime que 90 % des informations de l'« univers numérique » sont non structurées.

À elles seules, les données non structurées pertinentes en provenance des réseaux sociaux submergeront les entrepôts de données d'entreprise génériques. Plus vraisemblablement, les informations cruciales pour nombre d'analystes métiers seront ignorées, aboutissant au mieux à des résultats incomplets.

Concernant les types de données, il est important de rappeler que le modèle relationnel peut être désarçonné lorsqu'il tente de traiter des données non structurées. Aussi les bases de données relationnelles

sont-elles déclassées dès qu'il s'agit d'analyser des données XML hiérarchiques, des journaux Web non structurés et des graphiques point à point. Les systèmes analytiques spécialisés peuvent quant à eux gérer ces types de données car leur architecture les y autorise.

Tout en gérant l'accumulation des volumes et l'infinie variété (ou presque) des données, le moteur d'analyse spécialisé doit pouvoir jongler simultanément avec des charges de travail diverses et variées. Aucune tâche ne devrait mettre un système à genoux. Si l'administrateur de la base est occupé au transfert de nouvelles données depuis des sources externes alors que le système est sollicité de façon intensive par l'élaboration de rapports par lots, le système doit rester capable de traiter les requêtes ad hoc des communautés d'utilisateurs, sans aucun effort. La variation des charges de travail est un facteur déterminant de l'analyse décisionnelle du XXI<sup>e</sup> siècle.

Enfin, les analystes métiers doivent pouvoir utiliser des outils SQL standard pour exécuter leurs requêtes et ne pas être contraints de changer d'outils de visualisation pour présenter les données dans des tableaux de bord. En outre, les développeurs d'applications décisionnelles doivent pouvoir bénéficier des derniers perfectionnements en matière de programmation, comme le framework MAPReduce.

L'approche fondée sur une architecture spécialisée permet de mettre l'analyse décisionnelle à la portée de quiconque doit en disposer dans l'entreprise. Elle accepte le monde variable des types de données dans son intégralité tout en augmentant radicalement les volumes de données prêts à être traités. Et elle se charge de tout cela sans sacrifier les performances ni l'évolutivité.

Dans les pages qui suivent, vous apprendrez comment certaines entreprises résolvent les problèmes liés au décisionnel grâce à une technologie analytique spécialisée. ■■■

## Données non structurées : une véritable mine d'or



Le site Flickr héberge plus de **5 milliards d'images**

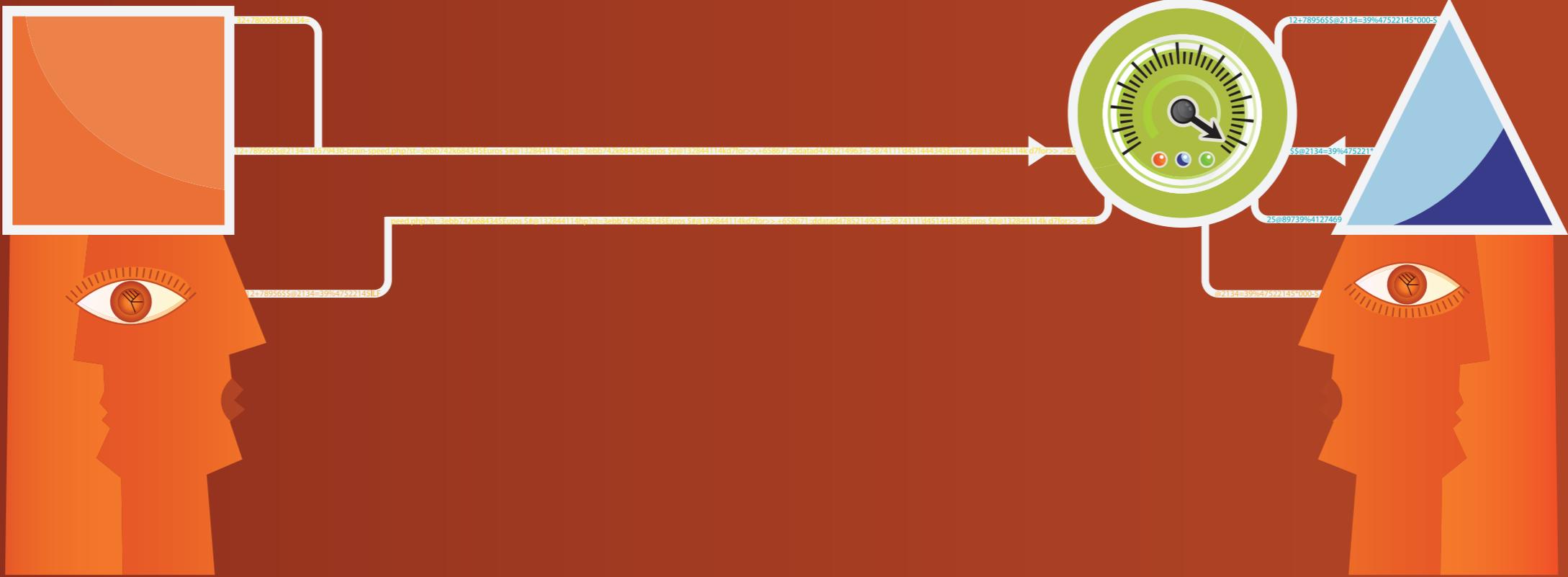
Plus de **30 milliards** d'éléments sont postés chaque mois sur Facebook

Chaque jour, **95 millions de tweets** sont publiés sur Twitter

Chaque minute, les utilisateurs de YouTube téléchargent **24 heures de vidéo**

LinkedIn inclut plus d'**un million** de profils d'entreprise

# L'analyse décisionnelle sur le terrain



**AUCUN SECTEUR D'ACTIVITÉ NE RESSEMBLE À UN AUTRE,** tout comme aucune entreprise ne ressemble à une autre dans un même secteur d'activité. Cependant, quel que soit le marché sur lequel elles évoluent, les entreprises sont confrontées aux deux mêmes problèmes interdépendants : que faire de toutes ces informations déversées dans les data centers à chaque seconde de chaque journée, et comment satisfaire les utilisateurs en nombre croissant qui souhaitent les analyser.

Dans le secteur de la santé, la généralisation des dossiers médicaux électroniques et l'analyse des informations sur les patients qui l'accompagne sont encouragées par des économies estimées à des dizaines de milliards d'euros par an<sup>17</sup>. Selon McKinsey & Company, si externaliser la chaîne logistique de fabrication permet de réaliser des économies, cela donne surtout aux cadres, analystes métiers et autres experts en prévisions les moyens de s'acquitter de

tâches essentielles comme l'acquisition, l'accès, la conservation et l'analyse de la quantité la plus importante d'informations possible. Et cela s'applique à tous les domaines imaginables : de la disponibilité des matériaux bruts au niveau des stocks chez un partenaire<sup>18</sup>. À l'instar de Wal-Mart, qui gère 267 millions de transactions quotidiennes<sup>19</sup>, les acteurs de la grande distribution réalisent pouvoir faire nettement pencher la balance en prenant leurs décisions de commercialisation sur la base d'informations pertinentes sur leur clientèle<sup>20</sup>.

Dans ces secteurs comme dans d'autres, la file d'attente des utilisateurs désireux de pouvoir analyser le flux de données entrantes ne cesse de s'allonger. Sans compter qu'une fois qu'ils y ont accès, ils exigent des temps de réponse de leurs requêtes aussi courts que ceux auxquels des moteurs de recherche comme Google ou Bing les ont habitués. Obtenir des informations et des réponses maintenant n'est pas

simplement une tendance ou une attente, mais bien une exigence liée à l'activité.

Quoi qu'il en soit, ce « maintenant » pour les simples mortels que nous sommes pourrait bien être insuffisant sur certains marchés car des vitesses de machine à machine s'imposent déjà dans certaines entreprises.

Selon *Wall Street & Technology*, la pression s'intensifie pour que les établissements de services financiers accélèrent les prises de décision, passant de la milliseconde, à la microseconde, puis à la nanoseconde. Cette analyse décisionnelle concerne les processus métiers dans le cadre des transactions financières exécutées par des ordinateurs puisque ceux-ci sont les seuls capables de fonctionner aussi rapidement que les données se déplacent. Les technologies nécessaires pour atteindre un tel niveau de fonctionnalités sont si onéreuses que seuls les acteurs majeurs du secteur peuvent continuer d'investir en matériel et en équipement de réseau pour régler le problème. À tous les autres, Larry Tabb, PDG du cabinet de conseil en technologie pour le secteur financier Tabb Group, prédit que « la prochaine étape consistera à prendre des décisions plus compliquées ». En d'autres termes, si les caisses de votre entreprise ne sont pas illimitées, vous devez faire preuve d'une intelligence accrue. Et pour rester dans la course, Larry Tabb vous conseille d'« élever les barrières de l'analyse décisionnelle<sup>21</sup> ».

Comme les services financiers, les entreprises de télécommunications sont confrontées à un raz-de-marée de données qu'elles doivent analyser en temps réel. Les principaux fournisseurs sur le marché européen ont vu leurs volumes de données exploser de 700 % uniquement pendant les trois années qui ont précédé 2010. En Amérique latine, les opérateurs de télécom ont dû accueillir un volume de données en



## Recevoir le message

**AUX ÉTATS-UNIS,** l'OPRA (Options Price Reporting Authority) fait partie de la petite quinzaine de sources de données financières parmi lesquelles figurent aussi le Consolidated Quote System, le New York Stock Exchange Open Book et le NASDAQ Trade Data Dissemination Service. Cette autorité fournit des détails sur les transactions effectuées et les options en cours, entre autres aspects essentiels sur les marchés quotidiens. Elle accumule les données à chaque seconde et la quantité d'informations ainsi produite croît à un rythme phénoménal.

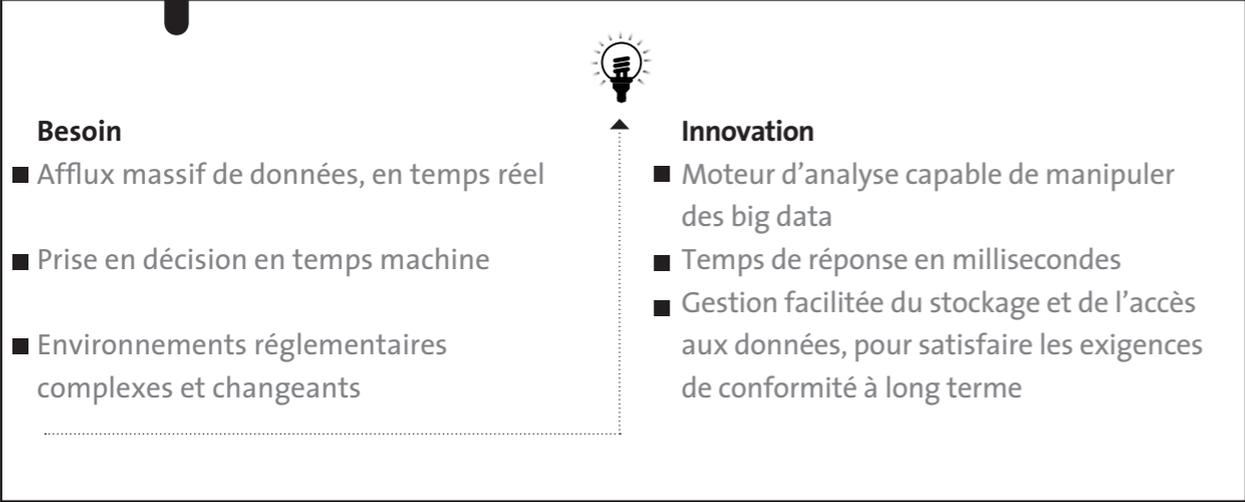
Pic de messages par seconde de l'OPRA, par année :

500	.....	1995
4,000	.....	2000
83,000	.....	2005
2,200,000	.....	2010

Les abonnés au service de l'OPRA doivent se préparer à prendre en charge une avalanche d'informations. La taille moyenne de chaque message par seconde étant de 120 octets, ce sont 264 Mo de données que l'OPRA déverse chaque seconde à elle seule. ■■■



### Innovation dans l'analyse décisionnelle : les services financiers



## Innovation dans l'analyse décisionnelle : la santé

<p><b>Besoin</b></p> <ul style="list-style-type: none"> <li>Faciliter la consolidation et l'analyse de grands volumes de données pour permettre aux organismes et fournisseurs de services de santé d'accélérer leurs recherches, d'améliorer les soins et de réduire les coûts.</li> <li>S'adapter à un environnement réglementaire en constante évolution.</li> </ul>	<p><b>Innovation</b></p> <ul style="list-style-type: none"> <li>Fonctions analytiques tirant parti de l'exploration des données (data mining), de l'analyse statistique, des prévisions, de la modélisation prédictive et des techniques d'optimisation pour guider vos décisions et vos actions stratégiques et opérationnelles.</li> <li>Technologie véritablement ouverte sur le monde changeant des normes sanitaires et dotée d'une solide expérience de l'intégration des processus métiers.</li> </ul>
---	---

entrée qui a doublé en quatre ans, passant de 5 à 10 téraoctets en moyenne. Il faut en outre s'attendre à une augmentation de ces volumes en raison du succès des smartphones. En effet, de 18,5 % en 2009, le trafic des données imputable à ces appareils sur les réseaux devrait bondir à 56 % en 2015, selon les prévisions<sup>22</sup>. De plus, comme sur le marché des services financiers, la vitesse et la masse des informations déversées dans les data centers des opérateurs de télécom sont propres à submerger un analyste humain. Pour répondre aux exigences de qualité de service, il est néanmoins nécessaire de prendre les décisions en temps réel, c'est pourquoi cela devra se faire par le biais d'une analyse décisionnelle intégrée dans des processus métiers gérés par ordinateur.

Face aux défis que posent les big data, certains mettent en garde les entreprises, les invitant à miser plus sur l'intelligence pour analyser les jeux de données, au nom du principe selon lequel « moins vaut mieux que plus ». Vous entendrez peut-être des fournisseurs relayer ce discours, d'autant plus

s'ils ne disposent pas d'une technologie capable de prendre en charge les volumes gigantesques de données. Dans ces conditions, on ne s'étonnera pas s'ils prétendent qu'il est « suffisant » d'appliquer des fonctions d'analyse à des jeux de données plus petits.

Une étude approfondie de ces arguments vous mènera au même constat que Hal Varian, économiste en chef chez Google : analyser une portion de données petite et aléatoire peut certes aboutir à des résultats pertinents. Mais pour obtenir un jeu de données véritablement aléatoire, il faut que la portion d'informations soit issue d'une énorme quantité de données. À défaut d'un réservoir de données suffisamment vaste, la validité de vos processus analytiques pourrait être remise en question. Pour résumer : les big data génèrent les meilleures données pertinentes<sup>23</sup>.

N'oubliez pas : si un fournisseur prétend que vous pouvez vous contenter de jeux de données restreints, c'est sans doute parce qu'il ne sait pas travailler sur des big data. 

# Comment transformer les big data en **atout de taille**



**D' participent déjà à l'élaboration de perspectives. Cela fait même quelque temps que des entreprises procèdent à des analyses décisionnelles basées sur des volumes gigantesques de données et en tirent un profit de taille.**

Prenons l'exemple de CoreLogic, filiale de First American, société figurant au classement Fortune 500 des établissements de services financiers. Désireuse d'innover dans le domaine des titres hypothécaires, CoreLogic s'est retrouvée confrontée à une double problématique : développer son activité et proposer des solutions répondant aux exigences toujours plus élevées de ses clients tout en faisant fonctionner une architecture technologique efficace. Elle devait aussi enrichir les fonctionnalités qu'elle proposait sur le Web et optimiser les fonctions existantes de ses applications d'analyse.

Problème : pour ses analyses décisionnelles, CoreLogic utilisait une base de données générique classique. Or, ce système ne pouvait pas traiter les charges de travail variables sur un jeu de données dont la taille se mesurait en téraoctets. Résultat : 20 % des requêtes des utilisateurs mettaient pas moins de cinq minutes à s'exécuter et 5 % s'interrompaient avant la fin par excès de données.

Désormais, CoreLogic dispose d'un moteur d'analyse spécialisé dont l'architecture repose sur une structure avancée de stockage en colonnes qui permet de traiter sans encombre les jeux de données très volumineux. Cela a décuplé les performances des applications décisionnelles, qui ont enregistré un gain de 800 %.

Pour CoreLogic, il ne s'agit pas seulement de quelques améliorations bluffantes. C'est une mise à niveau intégrale qui offre une expérience totalement nouvelle et différente aux utilisateurs, et plus

particulièrement aux utilisateurs expérimentés. Les technologies de télécommunication ayant évolué rapidement, une nouvelle ère s'est ouverte pour les opérateurs de télécom, dans un environnement toujours plus concurrentiel. D'autres défis se posent en outre au secteur, notamment le respect de la réglementation mondiale sur la conservation des données. La société Taiwan Mobile est concernée au même titre que ses concurrents.

Dans le cadre de leurs activités, les entreprises de télécommunications collectent d'énormes quantités de données brutes pour pouvoir facturer leurs abonnés et exécuter leurs transactions financières. La réglementation sur la conservation des données les contraint en outre à réaliser des enregistrements d'appels qu'elles doivent tenir à la disposition des services chargés d'en contrôler l'application. Les informations requises incluent la date, l'heure, la destination de l'appel, les numéros appelés et la durée de chaque appel. Elles doivent être conservées de 6 à 24 mois.

Taiwan Mobile est tenue de procéder aux enregistrements d'appels et au traitement de ces enregistrements dont la conservation est fixée à six mois. Avec plus de six millions d'abonnés mobiles, la quantité d'enregistrements d'appels et de transactions se chiffre à plusieurs milliards par jour. Taiwan Mobile estime que la capacité de stockage de toutes ces données pourrait facilement atteindre 10 téraoctets, ce qui submergerait le système existant, donc le ralentirait.

Pour atteindre ses objectifs, Taiwan Mobile a choisi de rejeter la solution de base de données générique. Cela lui permet de conserver des quantités importantes de données tout en bénéficiant de la capacité de compression de la technologie de stockage en colonnes. Elle peut ainsi réduire l'espace disque utilisé tout en analysant bien plus rapidement les données. ■■■



### Innovation Décisionnelle : les télécommunications

- Besoin**
- Analyse des clients – limiter les résiliations et augmenter le revenu moyen par abonné grâce à une rapidité accrue de l'analyse de grands volumes de données portant sur les profils des utilisateurs, les habitudes et les tendances.
  - Garantie de service – assurer la satisfaction des clients et remplir les conditions de niveau de service les plus strictes grâce à la possibilité d'intégrer et d'analyser en temps réel à la fois les données sur les utilisateurs et sur le réseau.
  - Optimisation de la capacité du réseau – optimiser les dépenses et l'utilisation du réseau en effectuant en temps réel la surveillance et les mesures de milliards d'événements sur le réseau, en descendant jusqu'au niveau du terminal.



- Innovation**
- Surveillance active des résiliations des clients – identification des schémas d'interruption et des opportunités pour améliorer la satisfaction.
  - Surveillance active des conditions du réseau et priorisation de son utilisation en fonction de la rentabilité des clients.
  - Unification de l'historique complet des données sur un client pour que le service d'assistance à la clientèle dispose de la vue la plus récente possible.
  - Possibilité de collecter, filtrer et agréger les données sur les clients pour procéder à une analyse historique par dimension : appels au support, historique de paiement, mises à niveau, habitudes d'utilisation et rentabilité globale.



## Moins d'espace, plus d'informations pertinentes

**E**N FORTE CROISSANCE, la société HealthTrans est l'un des principaux organismes américains de gestion de frais médicaux. Elle fournit aux employeurs, autres gestionnaires, courtiers, consultants, organismes de gestion de soins et autres acteurs du secteur, les outils et l'expertise clinique nécessaires pour gérer plus efficacement leurs programmes de gestion de dépenses pharmaceutiques. Elle gère plus de 100 téraoctets

de données et traite plus de 250 000 nouvelles transactions chaque jour.

Sa base de données traditionnelle tardait parfois jusqu'à 8 ou 10 heures pour produire des rapports. Elle a été remplacée par un système analytique spécialisé grâce auquel les rapports sont prêts en 2 à 3 minutes seulement. En plus, HealthTrans a réduit de 69 % ses besoins de stockage. ■■■

### Profil de structure en colonnes | HealthTrans

**2000**

Fondée en 2000

**15,3 millions**

15,3 millions de patients

**100**

100 téraoctets de données stockées

**69 %**

Besoins de stockage réduits de 69 %

**100 millions**

100 millions de demandes de remboursement par an

**\$4 milliards**

4 milliards de dollars US (2,9 milliards d'euros) de frais de médicaments sur ordonnance

**250,000+**

Plus de 250 000 transactions par jour

## La couverture du Cloud

**D**E NOMBREUX PETITS ÉDITEURS proposent des logiciels d'analyse décisionnelle en mode Cloud. Ils laissent généralement leurs clients exécuter des instances de leur application décisionnelle sur une infrastructure partagée. Les coûts de démarrage sont bas car, s'il est configurable, le logiciel se prête peu ou pas à une personnalisation (et lorsqu'il s'y prête, il faut payer le prix). Les frais engagés ne servent pas à acheter des logiciels ou du matériel, mais uniquement à payer les frais de fonctionnement, ce qui explique pourquoi tant d'entités au sein des entreprises s'abonnent à ce type de service sans passer par leur direction informatique.

Toutefois, bien qu'elle soit relativement bon marché et simple d'emploi, l'analyse décisionnelle sur le Cloud présente des inconvénients majeurs. Le plus évident concerne la sécurité des données. De nombreux prestataires de cloud computing ont inclus cette dernière dans leur solution, mais pas tous. Mais un problème encore plus sérieux se pose avec les big data. Certes,

l'analyse de gigaoctets de données dans le Cloud ne devrait pas être un souci, mais y charger des téraoctets d'informations à analyser peut se révéler une tâche très ardue. Tout aussi difficile en mode cloud : ouvrir le système à un nombre plus important d'utilisateurs. Et dans de nombreux cas, il est tout simplement impossible d'exécuter des requêtes ad hoc.

Parfaitement adapté aux petites entreprises qui ne sont pas de taille à gérer une infrastructure analytique, le décisionnel sur le Cloud leur permet de glaner des informations pertinentes sur l'activité et le marché. On peut aussi lui reconnaître l'avantage de soulager le service informatique des grosses entreprises dont une entité aurait momentanément besoin d'accéder à un système analytique si celui en place est dans l'incapacité de gérer la charge. Hormis dans ce cas, les offres analytiques sur le Cloud ne constituent pas un bon choix stratégique pour la plupart des entreprises d'envergure. ■■■

## Appliances : une envergure limitée

**U**N APPLIANCE d'entrepôt de données est une solution préconfigurée et pré-optimisée qui offre généralement une base de données, des outils de modélisation élémentaires et une technologie décisionnelle. L'ensemble peut entrer rapidement en service une fois les données chargées. Ce type d'appliance est généralement déployé comme solution tactique pour applications de mini-entrepôt de données.

La taille des jeux de données ne doit pas dépasser le

pétaoctet et les populations d'utilisateurs sont limitées. Il se trouvera cependant toujours un fournisseur pour affirmer que ses appliances peuvent prendre en charge de plus grands volumes de données et jusqu'à plusieurs milliers d'utilisateurs. Pour une administration simplifiée, les appliances doivent disposer d'une console unifiée. Leur retour sur investissement est rapide. Moins coûteuses, elles constituent parfois une solution idéale pour les petites à moyennes entreprises. ■■■

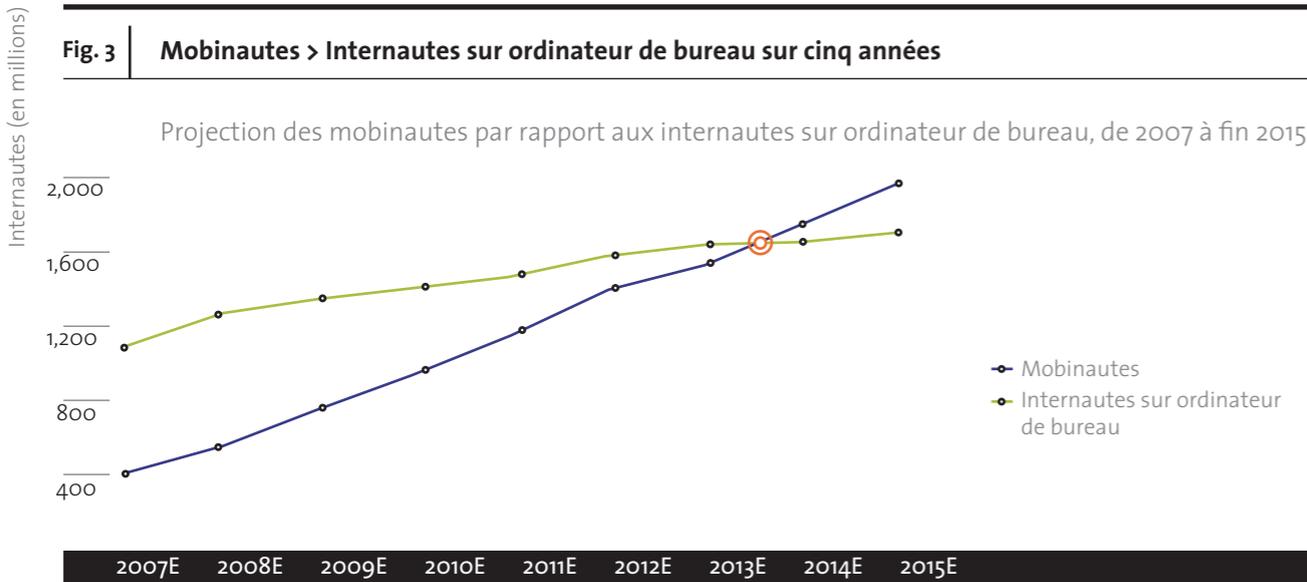


# Utilisateurs nomades

**D**ANS LE MONDE ENTIER, les services informatiques vont subir les assauts d'une armée d'utilisateurs mobiles réclamant l'accès à des fonctions analytiques (cf. figure 3). Selon Morgan Stanley, d'ici à fin 2014, le nombre des mobinautes aura dépassé celui des internautes qui surfent sur un ordinateur de bureau<sup>24</sup>. Cette tendance est confirmée par le Gartner Group qui estime que 80 % des entreprises auront confié une tablette à leurs employés d'ici à fin 2013<sup>25</sup>. Cela concerne bien sûr les effectifs sur le terrain dont la mission dépend d'applications décisionnelles et de la capacité à exécuter des

requêtes ad hoc. Si l'on considère que les utilisateurs finaux pourront vraisemblablement choisir leur plateforme de prédilection (iPad, iPhone, smartphone Android et autre appareil sous Windows), les services informatiques ne pourront pas s'offrir le luxe de vérifier qui utilise quoi pour accéder aux nouvelles applis décisionnelles. Cette tendance ne manquera pas d'accélérer la généralisation du décisionnel, se traduisant à court terme par un nouveau défi pour les équipes informatiques. ■■■

**Fig. 3** Mobinautes > Internautes sur ordinateur de bureau sur cinq années



Source: Morgan Stanley Research

# Plus d'utilisateurs, moins de maux de tête





# Une simple portion fait office d'échantillon

**D**E LA MÊME FAÇON qu'il est possible de tirer des conclusions pertinentes à partir de petites portions provenant d'un jeu de données volumineux, les études de cas présentées ici sont une petite portion extraite d'un très large ensemble d'entreprises ayant relevé les défis des big data et des populations grandissantes d'utilisateurs. À chaque fois, ces entreprises avaient auparavant fait l'expérience des bases de données SQL classiques génériques avant de constater que celles-ci ne répondaient pas à leurs exigences.

En optant pour une technologie spécialisée, non seulement les entreprises trouvent la solution à leurs problématiques en matière de décisionnel, mais elles se donnent les moyens de réussir la gestion de leur croissance future. Plutôt que d'envisager le flux de nouvelles données et l'arrivée de nouveaux utilisateurs métiers comme des problèmes, chacune d'entre elles peut désormais entrevoir de véritables opportunités de pénétrer de nouveaux marchés, développer de nouveaux produits ou services, réduire les coûts et optimiser le fonctionnement opérationnel global.

Ces opportunités donnent toujours lieu à un retour plus rapide et important sur les investissements dans les technologies analytiques spécialisées. ■■■

S'étant d'abord développé au niveau régional, l'établissement financier argentin Banco Macro a pris une dimension nationale avant de devenir l'une des principales banques du pays. La base de données SQL héritée de son passé régional commençait à céder sous la charge de travail imposée par une base d'utilisateurs passée de 6 à plus de 800 employés. L'exécution des requêtes les plus élémentaires pouvait prendre jusqu'à cinq minutes. Une fois la mise à niveau effectuée vers une architecture analytique spécialisée, ces mêmes requêtes élémentaires s'exécutaient en 20 à 30 secondes, malgré le doublement de la population des utilisateurs passée entre-temps à 1700 salariés et malgré des charges de travail bien plus lourdes sur des jeux de données plus de deux fois plus volumineux que la base de données SQL initiale. ■■■

d'externalisation. Plus de 12 000 utilisateurs exécutent jusqu'à 2000 requêtes ad hoc chaque jour sur son système analytique en colonnes, outre les rapports habituels. Malgré tout, les temps de réponses restent inférieurs à 20 secondes.

Le géant français des services financiers internationaux, BNP Paribas Securities Services, met les analyses décisionnelles à la portée de plus de 10 000 utilisateurs en leur donnant accès à environ 200 000 rapports. Le succès de ces fonctions analytiques spécialisées a conduit les utilisateurs à demander l'ajout de 10 000 à 20 000 rapports chaque mois. Les requêtes ad hoc atteignent quant à elles le nombre de 10 000 par mois. En outre, les clients utilisateurs finaux et les effectifs internes disposent dorénavant de données quasiment en temps réel dans l'ensemble des rapports et des requêtes.

**U**NE FOIS LE PROBLÈME DES BIG DATA RÉSOLU, la plupart des entreprises doivent encore relever l'autre partie du défi de l'analyse décisionnelle. Il est en effet tout aussi essentiel qu'elles prennent en charge une communauté d'utilisateurs grandissant sans cesse. Heureusement, elles peuvent y parvenir grâce à une architecture analytique spécialisée. Loin d'être uniquement théorique, cette approche se vérifie chaque jour sur le terrain.

Par exemple, Alvion Technologies, Inc. offre des services de gestion de données externalisés à quelques-uns des principaux fournisseurs mondiaux d'informations commerciales comme Axiom, Equifax, Experian et Dunn & Bradstreet. Les jeux de données de ses clients peuvent atteindre 190 millions d'enregistrements et 200 attributs. Au total, Alvion Technologies gère environ un téraoctet de données pour ses clients en tant que service

**Profil de structure en colonnes**

**BNP Paribas Securities Services**

**2001**

Filiale à 100 % de BNP Paribas Group SA, créée lors de la fusion de BNP et Paribas en 2001

**Première**

Première banque européenne de renom à se mettre en conformité avec Bâle II

**10,000 à 20,000**

10 000 à 20 000 nouveaux rapports chaque mois

**4552 milliards**

Actifs en conservation : 4552 milliards d'euros

**10,000**

10 000 requêtes ad hoc chaque mois

**10,000+**

Plus de 10 000 utilisateurs décisionnels

**734 milliards**

Actifs administrés : 734 milliards d'euros

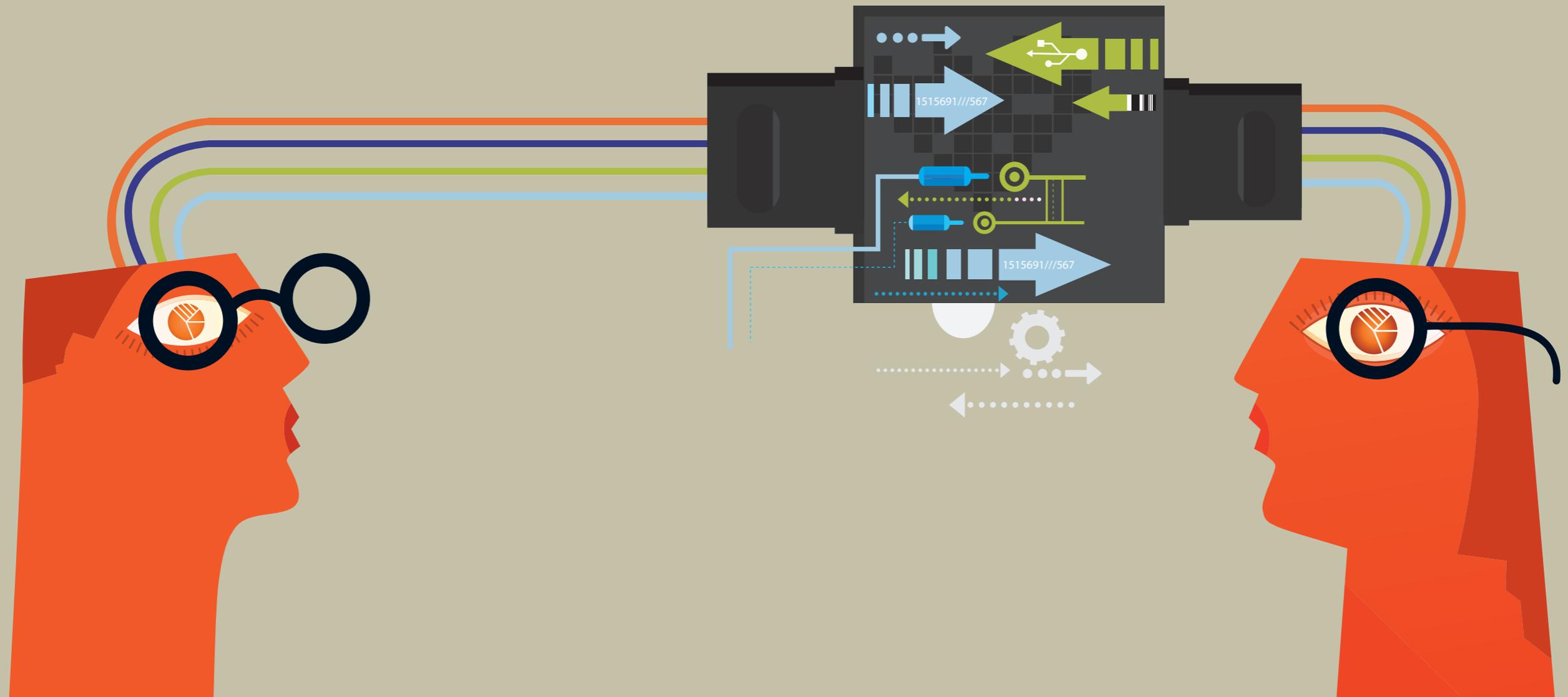
**200,000**

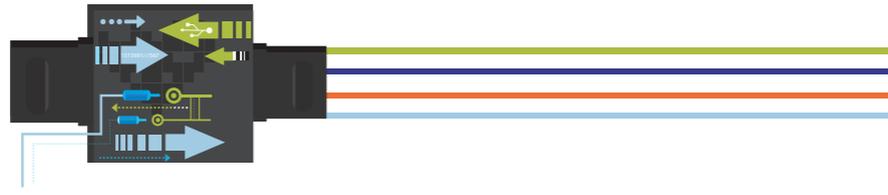
200 000 rapports

**Temps réel**

Réponses quasiment en temps réel

# Meilleures pratiques et retour sur investissement de l'analyse décisionnelle





En s'appuyant sur des exemples, la partie précédente nous a montré que de nombreuses entreprises amélioraient leurs résultats grâce aux applications analytiques. Les utilisateurs métiers constatent régulièrement des améliorations dans quasiment tous les domaines : du développement de produits disponibles plus vite sur le marché aux ventes dynamisées par des actions marketing ciblées, en passant par des actions plus rapides et précises sur les marchés financiers.

Il serait toutefois erroné d'affirmer que tous les projets décisionnels assurent un retour sur investissement positif car certains n'y parviennent tout simplement pas. Les raisons de l'échec sont généralement identifiées. Elles vont d'une planification insuffisante en amont aux performances et à l'évolutivité défailtantes du produit résultant souvent d'une mauvaise gestion de la démonstration de faisabilité<sup>26</sup>. Menée efficacement, cette dernière mettra en évidence aussi bien le potentiel en amont d'un projet analytique bien conçu, que nombre des faiblesses d'un projet mal préparé.



Le potentiel de la  
**démonstration** de  
faisabilité

**L**A DÉMONSTRATION DE FAISABILITÉ est la première étape sur le chemin qui mène au retour sur investissement. Pour parvenir à destination, il faut commencer, dans la mesure du possible, par conduire cette démonstration là même où sont administrées les ressources informatiques. Si vous suivez le fournisseur dans son propre labo de test pour y observer les résultats d'une démonstration de faisabilité menée par ses soins, vous y perdrez votre temps et votre argent (même si le fournisseur prend en charge vos frais de déplacement) et ne serez pas à l'abri des problèmes. En effet, le fournisseur peut charger une équipe d'experts d'optimiser manuellement l'environnement décisionnel pour le soumettre à des données et des requêtes préparées à l'avance. Il fera ainsi la preuve de l'efficacité de ses produits, mais vous n'aurez aucune garantie de la retrouver une fois la solution mise en production.

Ensuite, votre démonstration de faisabilité doit porter sur des données du monde réel, dans un environnement le plus sécurisé possible. Pour tirer des enseignements précieux de la démonstration de faisabilité d'une solution décisionnelle, n'acceptez pas les jeux de données de remplacement. Vous devez aussi veiller à ce que votre jeu de données soit utilisé dans son intégralité pour reproduire les conditions d'analyse à l'identique.

Bien sûr, la démonstration de faisabilité doit également porter sur la génération de vos rapports existants. La rapidité des temps de réponse devrait vous laisser bouche bée en comparaison des performances de vos applications actuelles. Dans le cas contraire, interrompez la démonstration et reprenez tout à zéro.

S'il est important d'améliorer les performances des rapports décisionnels existants, il est encore

plus essentiel d'inclure dans la démonstration de faisabilité les requêtes ad hoc que les utilisateurs souhaitent exécuter sur les données. Vous ne devez pas communiquer ces requêtes à l'avance. Le fournisseur doit pouvoir faire la démonstration de leur faisabilité dans des conditions d'exploitation.

Au cours de la démonstration de faisabilité du système, les performances testées doivent s'avérer très nettement supérieures à celles du système décisionnel existant. Vous devriez également constater qu'il est possible d'analyser une plus grande quantité et variété de données. Et même si vous n'utilisez pas encore les informations non structurées dans le cadre de vos opérations d'analyse décisionnelle, les inclure dans la démonstration de faisabilité vous permettra d'élargir le champ des possibles de votre entreprise, tout en testant la capacité actuelle du fournisseur à manipuler les types de données de plus en plus variées.

Enfin, ne vous fiez pas aux fournisseurs qui prétendent qu'une démonstration de faisabilité peut durer plusieurs semaines, voire un mois ou plus. Ce ne sont pas des concurrents sérieux. Une démonstration de faisabilité menée dans vos locaux, avec vos données, vos rapports et vos requêtes ad hoc ne devrait prendre que quelques jours, au pire une semaine, mais pas plus longtemps. ■■■



Voici les sept étapes définies par le magazine **CIO** pour déployer un système d'analyse décisionnelle d'entreprise<sup>27</sup> :

1. Assurez-vous que vos données sont saines.
2. Formez efficacement les utilisateurs.
3. Procédez rapidement au déploiement, puis effectuez les réglages au fur et à mesure. Ne passez pas trop de temps en amont sur le développement de rapports « parfaits » car les besoins évoluent en même temps que l'activité. Fournissez le plus rapidement possible des rapports efficaces, puis peaufinez-les.
4. Optez dès le début pour une approche intégrée lorsque vous construisez votre entrepôt de données. Assurez-vous de ne pas vous enfermer dans une stratégie de données impraticable à mesure que vous progressez.
5. Avant même de commencer, définissez clairement le retour sur investissement attendu. Dressez la liste des avantages spécifiques dont vous voulez bénéficier et confrontez-les à la réalité tous les trimestres ou tous les semestres.
6. Concentrez-vous sur les objectifs de l'entreprise.
7. N'achetez pas un logiciel décisionnel simplement parce que vous pensez en avoir besoin. Vous devez déployer votre système décisionnel en considérant que vous avez besoin de trouver des données chiffrées qui se trouvent quelque part, mais vous ne savez pas exactement où. ■■■



## Retour sur investissement prévisible

### Pour qui

- American Airlines
- State of Sao Paulo, Brazil
- Cell ©
- AOK Hessen
- HMV Japan

### Combien

- 1 M\$ (env. 750 000 euros) par an de **fraudes détectées**
- 100 M\$ (env. 75 millions d'euros) de **revenus non déclarés**
- 20 M\$ (env. 15 millions d'euros) économisés sur **un seul projet**
- 3,2 M\$ (env. 2,4 millions d'euros) de **fraudes détectées**
- 3 à 15 fois plus d'**achats en ligne**

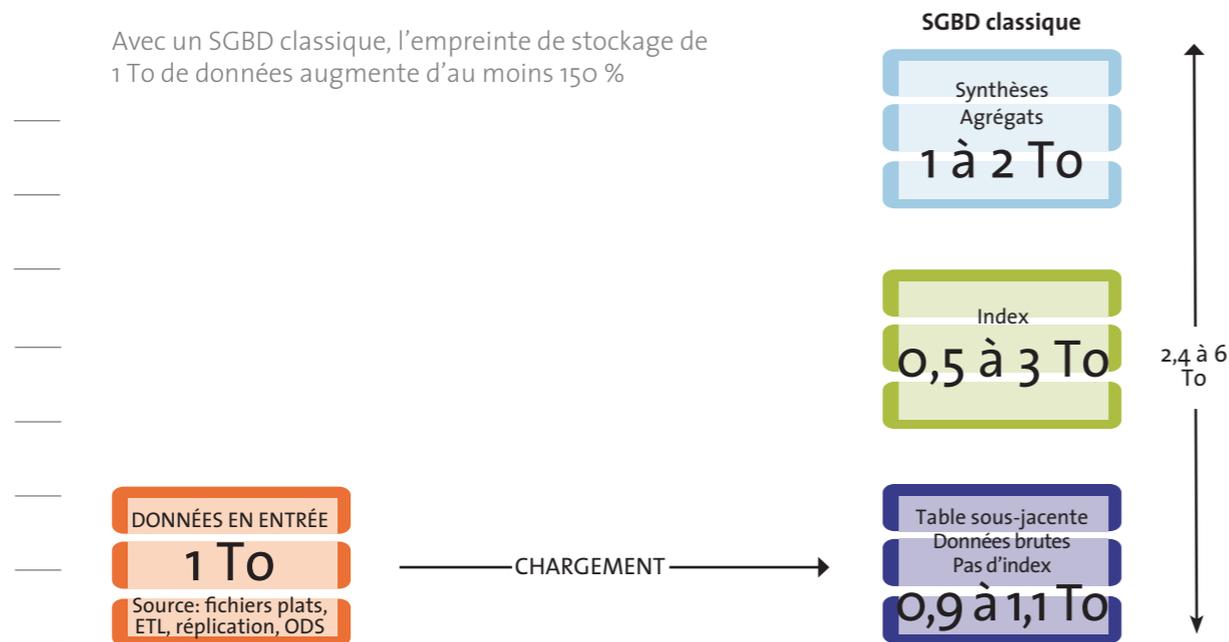
## Au-delà de la démonstration de faisabilité

**UNE FOIS LA FAISABILITÉ DÉMONTRÉE**, l'entreprise peut user de différentes tactiques éprouvées pour s'assurer le succès du projet décisionnel sur le long terme (cf. l'encart sur la page précédente « Les sept préceptes d'une analyse décisionnelle réussie »). Il faut avant tout que chaque personne impliquée dans le projet soit incluse dans la boucle pendant toute la durée de l'étape de déploiement. La communication doit être constante. Les intervenants

clés ne doivent jamais se trouver pris au dépourvu par des ratées au cours du processus.

Il existe un autre élément à prendre en considération au moment de définir le retour sur investissement potentiel d'une plate-forme logicielle analytique : la plate-forme matérielle. Cela peut sembler paradoxal, mais il faut bien que le logiciel s'exécute sur une architecture matérielle. Or, si vous le cantonnez à une seule configuration, vous limitez automatiquement les possibilités de retour sur investissement. Dans presque tous les cas, les équipements spécifiques sont plus onéreux à l'achat et probablement à l'entretien. Leurs composants coûtent plus cher. En outre, la gestion et l'administration d'un système sont par définition des activités coûteuses en raison des spécificités inhérentes au matériel. En clair, le coût total de possession vous empêchera de tirer pleinement profit de votre retour

Fig. 4 | Un modèle d'entrepôt de données traditionnel



sur investissement potentiel.

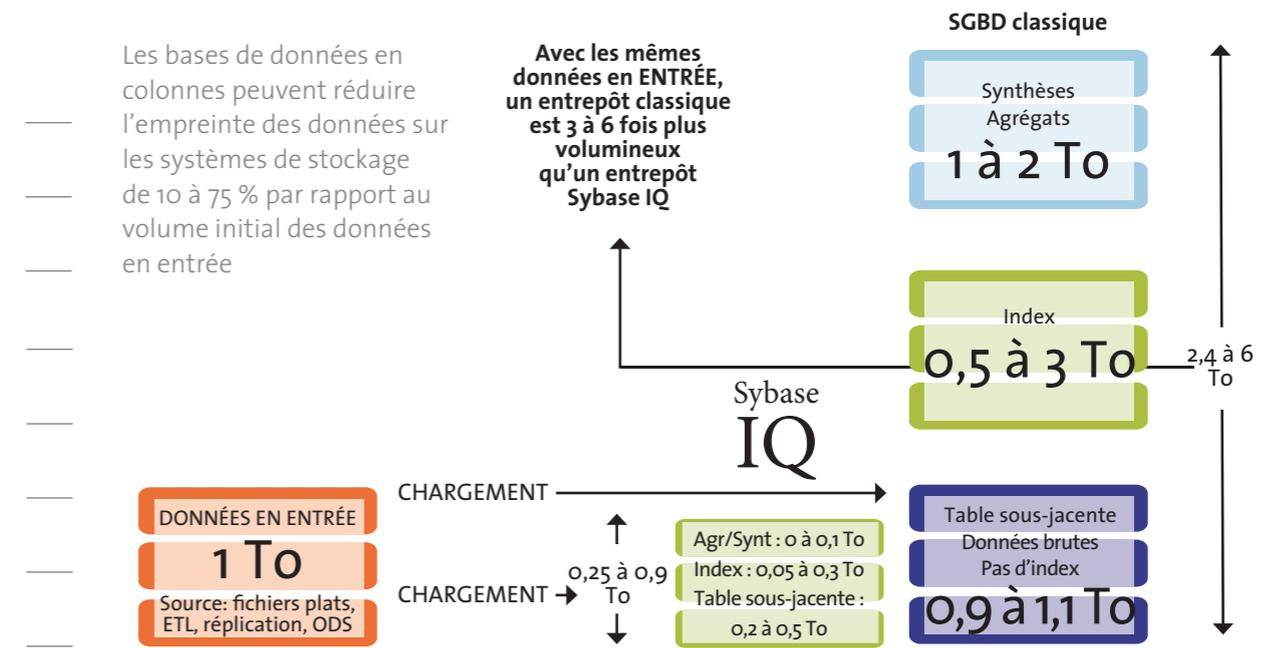
S'enfermer dans un choix matériel présente un autre inconvénient : vous devez attendre que le fournisseur procède à la mise à niveau de ses produits en fonction de son planning et non du vôtre. Pire, vous pourriez devoir renoncer à certaines avancées technologiques (en matière de microprocesseur, bus ou stockage, par exemple), simplement parce que votre fournisseur ne les inclut pas dans ses produits. Pour des performances et une évolutivité optimales, mettez toutes les chances de votre côté en ne vous enfermant pas dans le carcan d'un matériel unique.

Toujours concernant la plate-forme matérielle de votre système décisionnel et face à la réalité de la généralisation des big data, il faut tenir compte du rôle essentiel et évident du stockage dans le coût total de possession d'une application analytique, en particulier

s'il s'agit de prendre en charge des pétaoctets d'informations. Il faut bien stocker ces quantités de données quelque part. Selon le Gartner Group, « la croissance des données constitue pour les entreprises de taille importante le principal défi posé à l'infrastructure matérielle de leur data center »<sup>28</sup>. Dans un tel contexte, une architecture en colonnes spécialisée présente de réels avantages.

Comme l'illustre la figure 4, lorsqu'on charge des informations dans un entrepôt de données générique classique, l'encombrement qui en résulte est bien supérieur. Bien évidemment, on peut adjoindre à ce type d'entrepôt de données générique des méthodes de compression des informations. Toutefois, celles-ci sont totalement dépassées en comparaison des taux de compression des données de 10 à 75 % appliqués automatiquement par les bases de données en colonnes sans nécessiter l'ajout d'aucun module de complément (cf. figure 5).

Fig. 5 | Un modèle d'entrepôt de données traditionnel





# Calculer le retour sur investissement de l'analyse décisionnelle

IL EXISTE UNE ÉQUATION MATHÉMATIQUE permettant de calculer les avantages financiers dont les entreprises peuvent bénéficier en recourant à l'analyse décisionnelle (cf. figure 6). Les conclusions d'une étude portant sur 179 grosses sociétés adeptes de la prise de décision basée sur les données — un concept défini conjointement par des professeurs du Massachusetts Institute of Technology et de la Wharton School — ont montré chiffres à l'appui que les résultats et la productivité y étaient 5 à 6 % supérieurs<sup>29</sup>.

Des chercheurs de l'université du Texas (University of Texas) ont observé l'impact de l'analyse décisionnelle sur 150 entreprises du classement Fortune 1000, en particulier eu égard à leurs résultats financiers, leurs activités vis-à-vis de la clientèle et leur fonctionnement opérationnel<sup>30</sup>. Ils en ont conclu que le développement des produits justifiait à lui seul le déploiement d'un système analytique dans les entreprises de cette envergure. Selon l'étude, le chiffre d'affaires imputable à la capacité de l'entreprise à proposer des produits et services innovants augmente parallèlement à l'accessibilité des données et à la spécialisation de ces produits et services - et ses aspects bénéficient à leur tour de l'impact positif du décisionnel.

Fig. 6 Équation de calcul de l'impact de l'analyse décisionnelle sur l'entreprise

$$MV = \sum_{i=1}^n A_i + DDD \times A_i$$

À quel point cet impact est-il positif ? Les chercheurs ont déterminé qu'une entreprise avec un chiffre d'affaires de 16,8 milliards de dollars US (env. 12,5 milliards d'euros) encaisserait 64 millions de dollars US (48 millions d'euros) supplémentaires sur cinq ans en mettant un système d'analyse à la disposition d'un nombre « plus élevé de collaborateurs autorisés » susceptibles d'« exploiter les informations... pour mieux repérer les tendances et les schémas de la demande et pour améliorer les recommandations de prise de décision et de mise en correspondance des profils », tout cela au plus grand bénéfice des ventes. Ce type d'entreprise pourrait en outre y ajouter pour 14 millions de dollars US (env. 10,5 millions d'euros) de ventes à de nouveaux clients chaque année.

Le rapport de l'université du Texas mentionne également que dans les domaines opérationnels de l'utilisation d'actifs, des prévisions et de la planification et de la fourniture de produits et services dans le respect des délais, le recours systématique à l'analyse décisionnelle au sein de l'entreprise permet d'optimiser les résultats à tous les niveaux. Un usage plus général de l'analyse décisionnelle peut ainsi améliorer de 18,5 % les activités de planification et de prévision dans les conditions d'une entreprise type observée<sup>31</sup>.

Ces excellents chiffres s'expliquent différemment selon l'entreprise et le secteur d'activité. Toutefois, il suffit d'observer de plus près l'usage de l'analyse décisionnelle pour des applications en temps réel sur le marché des télécommunications pour constater qu'un environnement analytique à la pointe surclasse littéralement les systèmes d'aide à la décision du passé<sup>32</sup> (cf. figure 8).

Fig. 7 Tirer parti de l'analyse décisionnelle pour atteindre ses clients

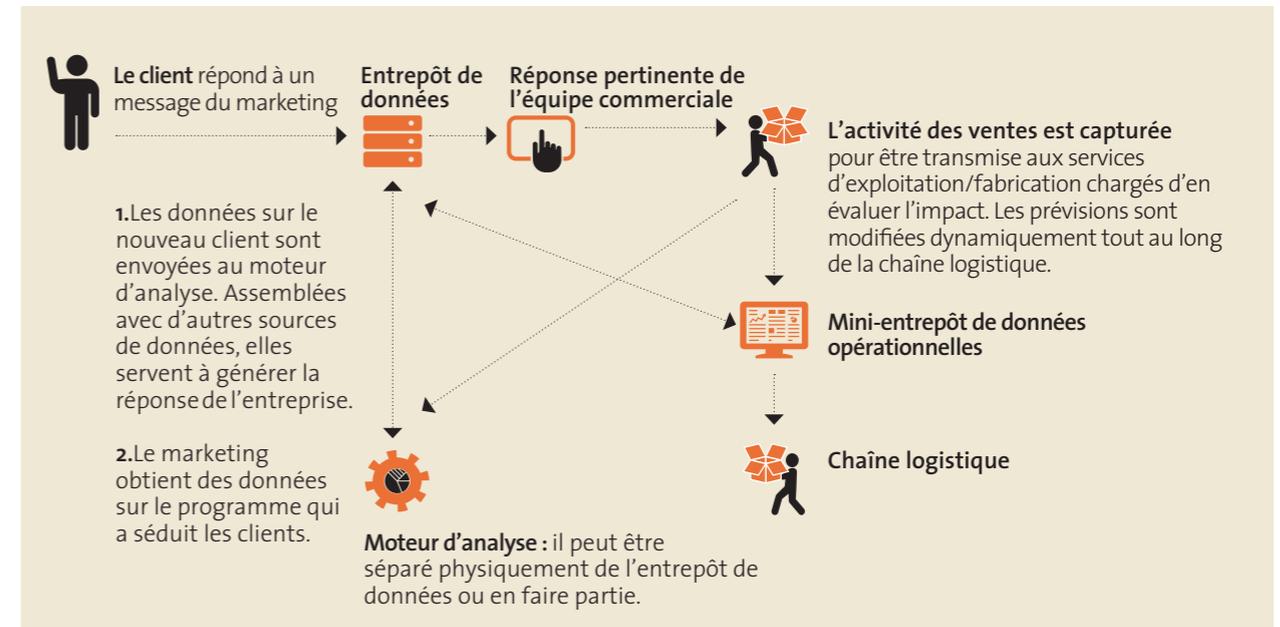
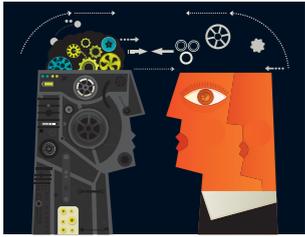


Fig. 8 Comparaison entre le décisionnel analytique du passé et celui en temps réel de nouvelle génération

	Infrastructure analytique du passé	Infrastructure analytique en temps réel de nouvelle génération
Coûts de stockage	Élevés	Bas
Type d'analyse	Hors ligne	En temps réel
Vitesse de chargement des données	Lente	Rapide
Durée de chargement des données	Longue	50 % plus courte en moyenne
Durée des tâches d'administration	Longue	60 % plus courte en moyenne
Temps de réponse des requêtes complexes	Heures ou jours	Minutes
Technique de compression des données	Pas mature	Compression des données supérieure de 40 à 50 % en moyenne
Coûts de support	Élevés	Bas



## De la recherche à la réalité

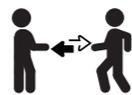
**IL EST RÉCONFORTANT DE CONSTATER** que des études exhaustives et factuelles menées par des universitaires et des analystes du secteur démontrent comment et pourquoi le retour sur investissement des applications analytiques est probable, voire assuré. Cela peut constituer l'un des pans de la justification globale d'investissements dans des projets décisionnels. Toutefois, rien n'est plus parlant que de vrais chiffres de retour sur investissement provenant de véritables entreprises.

Certes, de nombreuses entreprises préfèrent ne pas dévoiler à leurs concurrents les avantages spécifiques que leur procure l'analyse décisionnelle. Elles sont toutefois quelques-unes à révéler les profits et les économies dont elles bénéficient. Par exemple, l'État de Sao Paulo au Brésil s'est servi d'un système analytique spécialisé pour prendre de court les contribuables malhonnêtes en déterminant lesquels étaient susceptibles de ne pas acquitter leurs impôts. Cette opération a permis d'identifier directement des revenus non déclarés pour un montant total de 100 millions de dollars US (env. 75 millions d'euros). Autres avantages pour cette administration : outre qu'elle a réduit ses dépenses de stockage sur disque, elle génère dorénavant ses rapports à la vitesse fulgurante de quelques secondes contre plusieurs jours précédemment.

American Airlines a mis sur pied un programme analytique avancé pour déjouer une fraude : des escrocs contrefaisaient les billets d'avion au grand désespoir des passagers trompés et de la compagnie aérienne à laquelle l'arnaque coûtait très cher. Cette

opération lui a permis de réduire les pertes directement imputables à la vente des billets frauduleux d'un montant de 1 million de dollars US (env. 750 000 euros) en moyenne par an. Le service informatique de la compagnie a quant à lui réduit ses besoins de stockage de 70 %, un chiffre renversant.

Avec ses sept millions d'abonnés, l'opérateur de téléphonie mobile sud-africain Cell © disposait d'importants volumes de données qui augmentaient rapidement, mais était incapable d'en tirer parti pour détecter de nouvelles opportunités commerciales. En mettant en place une infrastructure analytique spécialisée, il a pu écourter les temps de génération des rapports d'analyse de huit heures à seulement quelques minutes, réduisant par là même ses coûts de 10 %. Enfin, dans le cadre d'un projet



### Omniprésence de l'analyse décisionnelle dans l'entreprise

- Modélisation financière
- Planification et développement de produits
- Analyse de fabrication et opérationnelle
- Acquisition de clients
- Évaluation de programmes de marketing
- Rétention des clients

d'installation de 80 stations de base à travers le pays, l'opérateur a pu économiser plus de 20 millions de dollars US (env. 15 millions d'euros).

Une fois son application analytique de détection des fraudes déployée, l'établissement allemand d'assurance santé AOK Hessen a rapidement repéré 179 factures frauduleuses représentant un montant total de 3,2 millions de dollars US (env. 2,4 millions d'euros) qu'il a pu se faire restituer. L'établissement est maintenant capable de cibler les fournisseurs qui facturent des taux sans rapport avec la réalité du marché.

HMV Japan commercialise de la musique et de la vidéo auprès du grand public. Le développement de programmes de marketing mieux ciblés grâce à des fonctions analytiques avancées lui a permis d'optimiser radicalement ses campagnes par email, avec une amélioration des taux de réponse de 50 à 70 %. Résultat : une démultiplication des achats de 3 à 15 fois selon la campagne et un bond du chiffre d'affaires par transaction de 150 à 300 %.

En Belgique, le Groupe Colruyt a obtenu le retour sur investissement de son déploiement analytique en trois ans. Grâce à sa structure en colonnes, le système a radicalement amélioré les performances de requêtes produisant des milliers de rapports et limité directement les besoins de stockage en réduisant jusqu'à 70 % les données et les index compressés. ■■■

## La passion des performances et des bénéfices

**C'EST NE SONT QUE QUELQUES-UNES** des nombreuses entreprises qui ont pu réaliser des bénéfices en s'appuyant sur un système d'analyse décisionnelle spécialisé. Présentes dans des pays et sur des marchés différents, elles partagent une

même passion pour la précision, les performances et, bien sûr, les bénéfices - c'est d'ailleurs ce qui les a motivées en premier lieu au moment d'investir.

Chaque entreprise a sa propre raison de déployer une architecture analytique spécialisée. Tirer parti des données pour s'ouvrir de nouvelles perspectives constitue déjà un avantage en soi, mais au-delà, l'entreprise peut avoir besoin d'augmenter sa population d'utilisateurs ou sa capacité en données, ou encore tout simplement d'améliorer ses performances grâce au système. Quoi qu'il en soit, toute entreprise attendra d'un environnement analytique avancé qu'il produise des effets positifs sur sa comptabilité.

D'une part, et comme nous l'avons vu dans les parties précédentes, ces attentes se fondent sur de nombreux précédents et sont confortées par des efforts de recherche objective considérables. Néanmoins, aucune plate-forme analytique ne ressemble à une autre. Les approches qui s'appuient sur une base de données générique classique achoppent sur les big data et sur l'arrivée d'un trop grand nombre d'utilisateurs concurrents sur le système. Elles sont lentes comparativement et peinent sous les charges de travail complexes.

D'autre part, une architecture analytique spécialisée bénéficiant des performances et de l'évolutivité d'une base de données en colonnes apporte de tels avantages chiffrables en termes de productivité — et dans de si nombreux domaines au sein de l'entreprise — que le retour sur investissement en devient presque secondaire, enfin presque.

Pendant, le retour sur investissement est déterminant ; il est même essentiel pour évaluer n'importe quel projet informatique d'envergure. Dans le cas de l'analyse décisionnelle, les chances d'un retour sur investissement sont là, à portée de main. Il ne dépend que de l'entreprise de la saisir. ■■■

- <sup>1</sup> **Structure Big Data Conference**, mars 2011, <http://event.gigaom.com/bigdata/>
- <sup>2</sup> **Gartner, Inc.** “Gartner Survey Shows Data Growth as the Largest Data Center Infrastructure Challenge,” novembre, 2010. <http://www.gartner.com/it/page.jsp?id=1460213>
- <sup>3</sup> **IDC**, The Digital Universe Decade: Are You Ready? mai 2010, p. 2.
- <sup>4</sup> **CED** in the History of Media Technology, <http://www.cedmagic.com/history/ibm-305-ramac.html>
- <sup>5</sup> **Everett M. Rogers et Judith K. Larsen**, Silicon Valley Fever, Basic Books, 1984, p. 252.
- <sup>6</sup> **Paul Russell**, 352 million computers sold worldwide in 2010, <http://www.discountvouchers.co.uk/news/85483535.html>
- <sup>7</sup> **Theodore Roszak**, The Cult of Information, Pantheon, 1986, p. 88.
- <sup>8</sup> **McKinsey Global Institute**, Big data: The next frontier for innovation, competition, and productivity, mai 2011, p. 2.
- <sup>9</sup> **IDC**, The Digital Universe Decade... p. 8.
- <sup>10</sup> **Statistiques issues de Wikipedia**, [http://en.wikipedia.org/wiki/Hard\\_disk\\_drive#cite\\_note-Mee-2](http://en.wikipedia.org/wiki/Hard_disk_drive#cite_note-Mee-2)
- <sup>11</sup> **Barua**, et. al., Measuring Business Impacts of Effective Data, Chapter One, septembre 2010, p. 2.
- <sup>12</sup> **Brochure**, 1995 to the First International Conference on Knowledge Discovery and Data Mining, <http://www.sigkdd.org/kdd1995/>
- <sup>13</sup> **Gartner, Inc.** Reveals Five Business Intelligence Predictions for 2009 and Beyond, <http://www.gartner.com/it/page.jsp?id=856714>
- <sup>14</sup> **Frank Hayes**, “The Story So Far,” Computerworld, 15 avril 2002.
- <sup>15</sup> **Colin White**, Is an Enterprise Data Warehouse Still Necessary for Business Intelligence? BI Research, 23 janvier 2011, <http://www.thevirtualcircle.com/2011/01/is-an-enterprise-data-warehouse-still-required-for-business-intelligence/>
- <sup>16</sup> **Eric Lai**, “Relational database pioneer says technology is obsolete,” Computerworld, 6 septembre 2007.
- <sup>17</sup> **Hillested**, et. al., “Can Electronic Medical Record Systems Transform Health Care? Potential Health Benefits, Savings, And Costs,” Health Affairs, 24, no. 5 (2005): 1103-1117
- <sup>18</sup> **Pande**, et. al., “Recapturing your supply chain data,” McKinsey Quarterly, mars 2006.
- <sup>19</sup> **David Bollier**, The Promise and Peril of Big Data, The Aspen Institute, 2010, p. 3.
- <sup>20</sup> **Rama Ramakrishnan**, “Impact of ‘Big Data’ on Retail,” CQuotient, 15 mai 2011, <http://blog.cquotient.com/?p=64>
- <sup>21</sup> **Entretien avec Larry Tabb** by Greg MacSweeney of Wall Street & Technology, 4 mai 2011.
- <sup>22</sup> **Ari Banerjee**, Addressing ‘Big Data’ Telecom requirements for Real-Time Analytics, Heavy Reading, mars 2011.
- <sup>23</sup> **Bollier**, The Promise and Peril of Big Data, The Aspen Institute, 2010, p. 15.
- <sup>24</sup> **Mary Meeker**, Internet Trends, MorganStanley, 12 avril 2010.
- <sup>25</sup> **Gartner, Inc.**, “Gartner Reveals Top Predictions for IT Organizations and Users for 2011 and Beyond,” 30 novembre 2010.
- <sup>26</sup> **Ferenc Mantfeld**, “Top 10 reasons why Business Intelligence Projects fail,” <http://www.seemoredata.com/en/entry.php?12-Top-10-reasons-why-Business-Intelligence-Projects-fail>
- <sup>27</sup> **Compilation de Ryan Mulcahy**, “Business Intelligence Definition and Solutions,” CIO, novembre 2007.
- <sup>28</sup> **Gartner, Inc.** “Gartner Survey Shows Data Growth as the Largest Data Center Infrastructure Challenge,” 1 novembre 2010. <http://www.gartner.com/it/page.jsp?id=1460213>
- <sup>29</sup> **Brynjolfsson**, et.al., Strength in Numbers: How Does Data-Drive Decisionmaking Affect Firm Performance?, Sloan Management School, 22 avril 2011, p. 6.
- <sup>30</sup> **Barua**, et. al., Measuring Business Impacts of Effective Data, septembre 2010.
- <sup>31</sup> **Ibid**, Chapter Three, p. 9.
- <sup>32</sup> **Ibid**, Banerjee, Addressing “Big Data”...

Nous espérons que vous avez apprécié et trouvé ce guide informatif sur la valeur de l'analyse décisionnelle, et plus particulièrement sur la transformation des volumes de données en informations utiles pour l'entreprise.

**Vous voulez en savoir plus?** Sybase est là pour vous aider. Nous offrons un large gamme de solutions de business intelligence, d'analyse en temps réel, de gestion des performances et d'entrepôt de données pour l'entreprise.

Nos technologies innovantes d'analyse accélèrent considérablement l'obtention de réponses à des questions complexes, et font émerger la valeur qui réside au sein des quantités massives de données de l'entreprise - qu'elles soient financières, opérationnelles, comportementales ou scientifiques.

Architecturées pour la flexibilité et l'efficacité, les technologies analytiques de Sybase aident plus de 2000 organisations dans le monde à gagner en pertinence, à améliorer leurs performances, à comprendre les tendances, à détecter la fraude, à gérer les risques et mieux satisfaire leurs clients.

**Pour plus d'informations**, retours d'expériences, livres blancs, ressources ou pour parler à un expert, visitez [www.sybase.com/analyze](http://www.sybase.com/analyze)

SYBASE, AN SAP COMPANY  
CORPORATE OFFICE  
ONE SYBASE DRIVE  
DUBLIN, CA 94568-7902 U.S.A.

1 800 8SYBASE  
COLLATERAL # L03337

[sybase.com](http://sybase.com)

Copyright © 2011 Sybase, an SAP Company. Tous droits réservés. Tous droits non publiés réservés conformément à la législation américaine en matière de copyright. Sybase et le logo Sybase sont des marques déposées de Sybase, Inc. ou de ses filiales. Le symbole ® désigne une marque déposée aux États-Unis d'Amérique. SAP et le logo SAP sont des marques ou des marques déposées de SAP AG en Allemagne et dans plusieurs autres pays. Toutes les autres marques appartiennent à leurs propriétaires respectifs. 01/11

**SYBASE**<sup>®</sup>  
An  Company