

# Une nouvelle intelligence pour une planète plus intelligente

*Les solutions analytiques IBM comme  
moteurs de l'innovation stratégique*



# **Une nouvelle intelligence pour une planète plus intelligente**

***Les solutions analytiques IBM comme  
moteurs de l'innovation stratégique***

**Pat Bates  
Mike Biere  
Rex Wiederanders  
Alan Meyer  
Bill Wong**



MC Press Online, LP  
Lewisville, TX 75077

## **Une nouvelle intelligence pour une planète plus intelligente**

*Les solutions analytiques IBM comme moteurs de l'innovation stratégique*

Pat Bates, Mike Biere, Rex Wiederanders, Alan Meyer, Bill Wong

Première impression—octobre 2009

© 2009 IBM Corporation. Tous droits réservés. Portions © MC Press Online, LP  
Portions © MC Press Online, LP

Nous nous sommes efforcés autant que possible de fournir des informations correctes. Toutefois, l'éditeur et l'auteur ne sauraient garantir l'exactitude de l'ouvrage et n'assument aucune responsabilité quant aux informations qu'il contient ou omet.

IBM est une marque d'International Business Machines Corporation aux États-Unis et/ou dans d'autres pays. Tous les autres noms de produit sont des marques commerciales ou protégées par les droits d'auteur de leurs fabricants respectifs.

Imprimé au Canada. Tous droits réservés.

Cette publication est protégée par le droit d'auteur ; à ce titre une autorisation doit être obtenue auprès de l'éditeur avant toute reproduction interdite, tout stockage dans un système d'archivage ou toute transmission sous quelque forme ou par quelque moyen que ce soit (électronique, mécanique, par photocopie, par enregistrement ou tout autre dispositif similaire).

MC Press offre des remises avantageuses sur cet ouvrage lors de commandes en grandes quantités ou lors d'offres spéciales, notamment des couvertures personnalisées et un contenu spécialement adapté à votre entreprise, vos objectifs de formation, votre orientation marketing et/ou à l'intérêt de votre marque.

Pour toute information concernant les autorisations ou les commandes spéciales, veuillez contacter :

MC Press  
Corporate Offices  
125 N. Woodland Trail  
Lewisville, TX 75077 USA

Pour toute information concernant les ventes et/ou le service client, veuillez contacter :

MC Press  
P.O. Box 4300  
Big Sandy, TX 75755-4300 USA

ISBN: 978-158347-086-2

# A propos des auteurs

**Pat Bates** est cadre supérieur au sein du groupe Worldwide Technical Sales Enablement, qui est axé sur l'entreposage de données et les technologies de logiciels d'analyse. Pat travaille dans ce domaine depuis 15 ans, dont les 8 dernières années chez IBM à divers postes ayant trait au développement logiciel, à l'encadrement, à la gestion des produits et la facilitation des ventes techniques. Il est intervenu dans de nombreuses conférences du secteur et a contribué à plusieurs articles et ouvrages dans les domaines de l'entreposage de données et de la business intelligence (veille économique).

**Mike Biere** est cadre supérieur dans l'équipe internationale IBM chargée de l'entreposage de données et de la Business Intelligence. Il a occupé de nombreux postes au sein d'IBM en tant que spécialiste technique de base de données, d'entreposage de données et de BI depuis 1978. Il a eu trois carrières IBM distinctes avec des missions à l'extérieur, en tant que consultant en entreposage de données / en business intelligence et en tant que directeur de gestion de produits pour Cognos. Il intervient régulièrement comme participant / conférencier lors d'événements tels que celui du groupe IDUG et est l'auteur de *Business Intelligence for the Enterprise*.

**Alan Meyer** est directeur marketing de l'activité entreposage de données au sein de l'organisation InfoSphere d'IBM. Il est présent dans la communauté de la BI depuis le milieu des années 1980 et a occupé des postes aussi bien techniques que marketing en entreposage de données et BI. Ses expériences antérieures recouvrent un large champ de compétences en maintenance du matériel, programmation d'applications, programmation systèmes et conseil. Il intervient régulièrement lors de manifestations du secteur et a écrit de nombreux documents et articles au sujet de l'innovation en lien avec l'information et détient plusieurs brevets logiciels.

**Rex Wiederanders** est responsable du groupe Worldwide Technical Sales Enablement, dont il fait partie depuis 12 ans. Rex a occupé différents postes au sein d'IBM depuis 1981, y compris en ventes, technico-commercial, développement et spécialités hors régions. Consacrant l'essentiel de ses efforts sur les logiciels depuis les 20 dernières années, Rex intervient régulièrement dans de nombreux salons professionnels et groupes d'utilisateurs, où son style de présentation unique est fort apprécié.

**Bill Wong** est responsable en gestion de l'information et chargé de concevoir des solutions de transformation induite par l'information. Il a participé aux offres de bases de données et d'entrepôt de données d'IBM depuis leur création, a occupé des postes en développement, marketing et ventes, et a passé plusieurs années à aider les entreprises à mettre en œuvre des solutions dans le secteur des services financiers, de la vente au détail et des secteurs des télécommunications. Il intervient régulièrement lors de manifestations du secteur, et les ouvrages qu'il a précédemment publiés portaient sur la technologie de base de données, les solutions de business intelligence dans le secteur des sciences de la vie et l'optimisation de l'activité grâce à des informations éprouvées.

# Remerciements

Nous tenons à exprimer notre gratitude et nos remerciements les plus sincères aux personnes suivantes pour leur soutien, leurs conseils et les contributions qu'elles ont apportés à ce premier ouvrage :

Kevin Sy, Aliye Ergulen, Catherine Haddad, Gary O'Connell, Stephanie Best, Natasha Engan, Susan Visser, Rick Miller, Lynn Jonas, Stephanie Clark, Drew Friedrich, Caryn Meyers, Meg Dussault, Guenter Sauter, John Rollins, Adam Gartenberg, Bill O'Connell, Karl Freund, Maria Winans, Inhi Cho, Greg Lotko, Arvind Krishna, Ambuj Goyal et Steve Mills



# Sommaire

<b>Avant-propos</b> . . . . .	<b>xi</b>
<b>Chapitre 1 : Présentation de notre concept Smarter Planet, pour une planète plus intelligente</b> . . . . .	<b>1</b>
Comprendre notre concept Smarter Planet, pour une planète plus intelligente . . . . .	4
Pourquoi développer une nouvelle intelligence ?	
Pourquoi maintenant ? . . . . .	9
Découverte de la nouvelle intelligence . . . . .	11
<b>Chapitre 2 : Présentation de l'analyse stratégique</b> . . . . .	<b>13</b>
Interactions entre utilisateur et informations . . . . .	16
Infrastructure d'analyse dédiée à la nouvelle intelligence . . . . .	28
IBM InfoSphere Streams . . . . .	32
Traitement analytique en ligne (OLAP) . . . . .	36
Solutions IBM InfoSphere Entity Analytics . . . . .	38
Exploration de données . . . . .	39
Analyse textuelle . . . . .	44
Gestion de contenus . . . . .	46
La couche d'infrastructure . . . . .	49
Le point sur la nouvelle intelligence . . . . .	53
<b>Chapitre 3 : Business Intelligence et gestion de la performance.</b> . . . . .	<b>55</b>
Les trois questions essentielles au maintien de la performance des entreprises . . . . .	60
Outils de BI / gestion de la performance : les principes de sélection pour les entreprises . . . . .	62
IBM Cognos 8 : une nouvelle suite au service d'une planète plus intelligente . . . . .	63
Collaboration et communication efficaces . . . . .	71
Le dilemme : des données structurées aux données non-structurées . . . . .	72
La segmentation des utilisateurs au service de la Business Intelligence . . . . .	75
Gestion des systèmes et accords de niveau de service . . . . .	77



Standardisation de la Business Intelligence . . . . .	78
Performances et évolutivité. . . . .	79
Synthèse . . . . .	80
<b>Chapitre 4 : Conception de l'infrastructure informatique . . . . .</b>	<b>83</b>
Pouvez-vous compter sur vos informations ? . . . . .	83
Comment accroître la confiance des utilisateurs envers vos informations ? . . . . .	84
Compréhension des informations. . . . .	85
Fiabilité des informations. . . . .	94
Accès à toutes les informations . . . . .	98
Automatisation de l'infrastructure informatique . . . . .	101
Regroupement des données aux fins d'analyse . . . . .	102
Master Data Management (gestion de données de référence). . . . .	105
<b>Chapter 5 : Le programme Information Agenda . . . . .</b>	<b>109</b>
Stratégie de gestion des données . . . . .	111
Concrétisation des objectifs . . . . .	115
Conclusion. . . . .	116

# Avant-propos

Bienvenue et merci de prendre le temps de comprendre l'initiative « IBM Smarter Planet », centrée sur la nouvelle intelligence.

Le monde dans lequel nous évoluons est sans cesse plus équipé en technologie, interconnecté et intelligent, et l'information est au cœur de cette révolution. Les entreprises qui tirent parti de cette nouvelle mine d'informations sont à même de prendre des décisions plus avisées et parviennent à se hisser vers le haut. Elles gèrent de gros volumes d'informations en temps réel, intégrant des modèles analytiques et prédictifs, collectant et partageant systématiquement des informations sur toute la chaîne de valeur, et accélérant le retour sur investissement en offrant des informations éprouvées, exactes et opportunes aux décideurs pertinents. En bref, elles découvrent un nouveau type d'intelligence.

De nos jours, les entreprises intelligentes ne se contentent plus de regarder vers le passé ; elles exploitent au contraire cette mine d'informations pour se projeter vers l'avenir. Cette nouvelle intelligence ne se contente pas de vous procurer un aperçu de vos opérations courantes : elle vous offre une perspective sur l'avenir.

Elle vous donne une idée de ce qui vous attend dans un futur immédiat et un peu plus éloigné. Les outils d'analyse et de reporting décortiquent les données afin de cristalliser les tendances, les schémas et les anomalies qui apportent un précieux éclairage sur le marché pour vous aider à faire des choix plus judicieux.

Partenaire de votre réussite, IBM a réuni ses meilleurs experts en la matière pour vous présenter ce guide d'introduction afin de vous accompagner et de soutenir vos efforts de transformation induite par l'information.

*Arvind Krishna*

*Arvind Krishna*

*Directeur général, Gestion de l'information*

*Groupe logiciels IBM*



# 1

## **Présentation de notre concept Smarter Planet, pour une planète plus intelligente**

**L**es dirigeants du monde entier polarisent plus que jamais leur attention sur les implications économiques, sociales et environnementales de l'intégration à l'échelle mondiale, où les accords de libre-échange, Internet et la mondialisation abolissent les distances et rendent notre monde plus accessible et aussi plus intelligent. Ce phénomène est en définitive susceptible d'avoir une incidence plus profonde sur notre société, les entreprises et la vie de chaque individu ; notre planète devient elle aussi de plus en plus intelligente. L'initiative IBM Smart Planet se concentre sur l'infrastructure mondiale ; ces systèmes et processus qui permettent aux marchandises d'être développées, fabriquées, achetées et vendues, et ces services essentiels qui fournissent aussi bien de l'électricité et des transactions financières que des soins de santé efficaces, pour n'en citer que quelques-uns, et qui touchent directement des milliards de vies.

L'infrastructure mondiale gagne en « intelligence ». Considérez les faits suivants :

- On dénombre plus d'1 milliard de transistors par être humain
- On estime à 2 milliards le nombre de gens à être prochainement connectés à Internet
- Parallèlement, nous approchons du trillion d'objets connectés, constituant « l'Internet des choses »

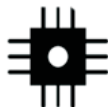
- Le nombre d'abonnements de téléphonie mobile dans le monde a dépassé les 3,3 milliards : une personne sur deux possède un téléphone portable sur notre planète
- Plus de 30 milliards de balises RFID (identification par radiofréquence) sont produites dans le monde, intégrées dans les produits, les passeports, les bâtiments et même les animaux
- Des centaines de satellites sont en orbite autour de la Terre, générant des téraoctets de données chaque jour

Et il ne s'agit pas uniquement d'une question de connectivité permanente. Pour la première fois, des ordinateurs massivement puissants peuvent être déployés à moindre coût pour permettre le traitement, la modélisation, la prévision et l'analyse de n'importe quelle charge de travail ou tâche. L'informatique dématérialisée (Cloud computing) apparaît comme un moyen de connecter et d'approvisionner la palette foisonnante de périphériques utilisateurs, capteurs et mécanismes d'accès dotés de systèmes dorsaux performants et hautement évolutifs. Le modèle informatique des années 1980 a fait place à un nouveau paradigme, fondé sur l'ouverture, les réseaux, les nouvelles technologies de pointe et l'intégration de l'intelligence numérique dans la trame de nos activités personnelles et professionnelles au quotidien.

Avec l'accroissement de l'intelligence de ces systèmes, trois caractéristiques distinctes les définiront de plus en plus (Fig. 1.1). Ils seront :

- *Equipés en technologie* : toute activité ou tout processus peut désormais être mesuré(e), mieux compris(e), modélisé(e) et amélioré(e) pour conférer une meilleure visibilité.
- *Interconnectés* : en puisant dans l'intelligence collective de toute la chaîne de valeur grâce à la connexion de systèmes entiers, le monde est à même de devenir plus auto-réglementé, optimisé et efficace.
- *Intelligents* : chaque information provenant de ce monde des périphériques intelligents peut se traduire par une valeur incrémentale, en permettant d'automatiser le traitement de certaines actions et avec beaucoup plus de certitude.

Les entreprises sont désormais en mesure de prendre des décisions plus rapidement et en étant mieux informées pour obtenir de meilleurs résultats commerciaux, car leurs systèmes peuvent être équipés en technologie, interconnectés et intelligents.



**EQUIPE EN TECHNOLOGIE**

Nous sommes désormais capables de mesurer, détecter et contrôler l'état de quasiment tout.



**INTERCONNECTE**

Les gens, les systèmes et les objets communiquent et interagissent les uns avec les autres de manière totalement innovante.



**INTELLIGENT**

Nous faisons preuve d'une grande réactivité face aux changements, rapidement et avec précision, et parvenons à de meilleurs résultats en anticipant et en optimisant nos ressources pour l'avenir.

*Figure 1.1 : le monde est sans cesse plus équipé en technologie, interconnecté et intelligent*

Pour la première fois dans l'histoire, presque tous les éléments de notre quotidien peuvent être traités d'un point de vue numérique et interconnectés. La planète devient chaque jour un peu plus intelligente, un peu plus équipée en technologie et un peu plus interconnectée. Ces évolutions s'accompagnent d'opportunités fascinantes pour la société dans son ensemble et pour chaque entreprise, institution et individu. Quand autant de technologie et de développement de réseau sont disponibles à profusion, qu'est-ce qui peut encore échapper à la technologie intelligente ? Quel service ne fourniriez-vous pas ? Quel élément ne connecteriez-vous pas ? Quelle information n'exploiteriez-vous pas pour posséder de nouvelles connaissances ?

La réponse est la suivante : « Rien n'est impossible ». Grâce aux technologies modernes, vous - ou vos concurrents - parviendrez à tout réaliser, et ce parce que vous le pouvez. Mais la raison impérieuse pour laquelle nous allons tous commencer à transformer nos systèmes, nos modes de fonctionnement, nos entreprises et nos vies personnelles pour tirer profit d'un monde plus intelligent ne tient pas au seul fait que nous sommes en mesure de le faire ; c'est parce que nous devons le faire.

Grâce au concept Smarter Planet, les entreprises excellent dans trois domaines (Fig. 1.2) :

- *Mettre l'accent sur la valeur* : les entreprises intelligentes parviennent à accomplir davantage avec moins de ressources ; elles se concentrent sur leur cœur de métier et tirent parti de leurs relations pour pérenniser une solidité financière tout en revisitant et revalidant sans cesse leurs objectifs fondamentaux afin de rester suffisamment souples pour évoluer au rythme de leurs clients.

- *Tirer parti des opportunités* : les entreprises intelligentes savent comment gagner des parts de marché, développer les capacités futures en préservant et en acquérant des talents, et figurent parmi les plus performantes dans leurs secteurs, n'hésitant pas à faire preuve d'audace lorsque cela est nécessaire.
- *Agir avec célérité* : les entreprises intelligentes ont à leur tête des dirigeants qui obtiennent les informations nécessaires pour mettre en œuvre, définir et communiquer la stratégie de l'entreprise, et qui permettent aux responsables des diverses branches d'activité de s'adapter aux changements et d'évoluer rapidement.



Figure 1.2 : les entreprises intelligentes doivent se concentrer sur la valeur, les opportunités et la célérité

## Comprendre notre concept Smarter Planet, pour une planète plus intelligente

Dans ce nouveau monde, où chaque être humain, entreprise, organisation, ville, nation, système naturel et anthropique devient interconnecté(e), équipé(e) en technologie, et est littéralement rendu(e) plus intelligent(e), nous estimons qu'il est nécessaire d'examiner avec attention les problématiques suivantes (Fig. 1.3).

Pour que nous puissions donner un sens à ce monde en pleine mutation, nous devons examiner quatre problématiques essentielles.

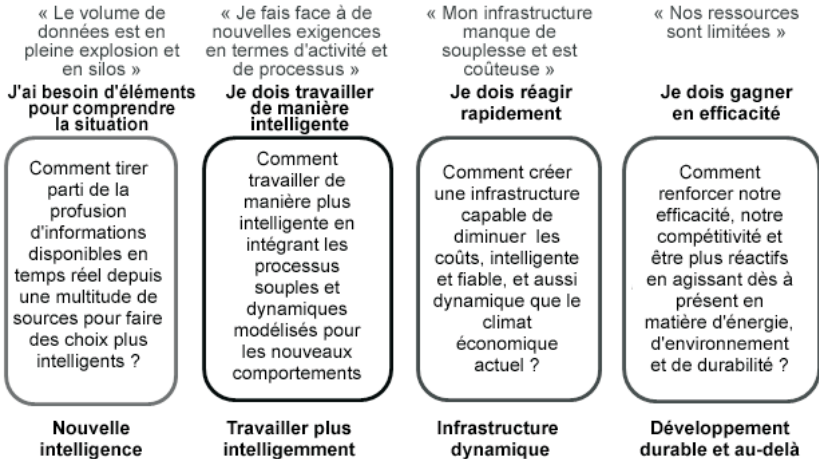


Figure 1.3 : quatre problématiques essentielles qui orientent l'activité

Abordons maintenant chacune de ces problématiques en détail. Pourquoi sont-elles importantes ? A quelles pressions sommes-nous confrontés ? Vers où nous tourner pour trouver des solutions ? Quelles solutions ingénieuses ont déjà mis en œuvre les entreprises pour résoudre leurs problèmes ?

## Nouvelle intelligence

Les entreprises intelligentes tirent parti de la mine d'informations disponibles pour prendre des décisions plus avisées. Les solutions de nouvelle intelligence présentent des avantages sur l'une ou plusieurs des dimensions suivantes :

- *Gestion des informations* : gestion et utilisation efficaces du volume et de la variété en pleine explosion des informations historiques et en temps réel
- *Fonction prédictive* : identification et traitement proactif des opportunités et des menaces grâce à l'analyse prédictive en temps réel
- *Développement de la chaîne de valeur* : des informations omniprésentes recueillies à partir d'une multitude de sources dans la chaîne de valeur et utilisables par tous ceux qui en ont besoin
- *Rapidité d'exécution* : accélérer le retour sur investissement en offrant des informations fiables, éprouvées et au moment opportun aux décideurs pertinents à tous les niveaux de l'entreprise



**Utiliser la nouvelle intelligence : des décisions concrètes en temps réel**  
Conclusions des perspectives Smarter Planet

- Les problèmes majeurs auxquels les responsables informatiques et l'entreprise sont confrontés sont le gaspillage, l'inexactitude et la quantité d'opportunités manquées à cause d'informations non maîtrisées.
- Les DSI ont l'occasion de prouver que les informations en temps réel peuvent être gérées plus efficacement pour créer des business models innovants et performants grâce à la nouvelle intelligence.
- Les DSI doivent puiser dans l'intelligence de toute leur chaîne de valeur, corréler les connaissances et anticiper les possibilités et les menaces.

Figure 1.4 : problèmes liés à la nouvelle intelligence

## **Travailler plus intelligemment**

Travailler plus intelligemment consiste à mettre en œuvre de nouveaux niveaux d'intelligence dans la manière dont les personnes, les entreprises et les systèmes anthropiques interagissent. Chaque interaction représente une chance de faire les choses mieux, plus efficacement et de façon productive par le biais de business models agiles et souples.

De plus en plus, les entreprises comptent sur leurs employés, clients, partenaires et fournisseurs pour améliorer leurs collaborations et interactions. Les processus doivent s'adapter dynamiquement à l'évolution des conditions économiques pour limiter les risques et entraîner des bénéfices. Les éléments fondamentaux pour une planète plus intelligente existent déjà, et travailler plus intelligemment est aujourd'hui une réalité.

Les attributs essentiels pour travailler plus intelligemment sont les suivants :

- *Des business models agiles* aident les entreprises à changer rapidement de cap afin de refléter les réalités économiques.
- *La collaboration* aide les différents protagonistes à étendre les relations internes et externes en vue d'accroître l'expertise et d'accéder à de nouvelles informations et ressources en dehors de leurs groupes de travail habituels.
- *Des processus métier dynamiques* aident les entreprises à accéder à des informations éprouvées, réutiliser les actifs et réduire les coûts en automatisant les processus manuels.

- *Des architectures orientées services intelligentes (SOA) convertissent des applications isolées et inefficaces en services réutilisables.*

**Travailler plus intelligemment : s'adapter aux changements, donner aux collaborateurs les moyens d'agir, augmenter les profits**  
Conclusions des perspectives Smarter Planet

- Les entreprises faisant usage de business models flexibles, de processus métier dynamiques et d'environnements de travail collaboratifs seront particulièrement bien placées pour être les meilleures.
- De la personnalisation, de la co-création et des compétences à la fois commerciales et informatiques sont nécessaires pour toucher les nouveaux clients connectés.
- Ces attributs permettent aux entreprises de fournir un environnement utilisateur de qualité aux clients, fournisseurs, employés et partenaires.
- La capacité de travailler plus intelligemment et de s'adapter à l'évolution des conditions économiques sont essentielles pour limiter les risques et accroître les bénéfices.

*Figure 1.5 : travailler plus intelligemment*

## **Infrastructure dynamique**

Une infrastructure dynamique est conçue pour aider à transformer les actifs physiques et numériques en services à plus forte valeur. Une infrastructure dynamique est hautement optimisée pour parvenir à de meilleurs résultats grâce à une gestion améliorée, et tire parti des nouvelles technologies et stratégies en vue de réduire les coûts et d'offrir des services commerciaux et informatiques de premier ordre avec souplesse et célérité.

Une infrastructure dynamique confère trois avantages distincts pour toute l'entreprise, tout en posant également les bases pour l'avenir :

- *Améliorer le service* : en interne comme en externe, les clients et les employés attendent un service de premier ordre, non seulement en termes de haute disponibilité et de qualité des services existants, mais aussi pour répondre à leurs attentes croissantes pour un accès en temps réel et dynamique à de nouveaux services innovants.
- *Réduire les coûts* : la réduction des coûts fonctionnels est certes importante, mais une infrastructure dynamique peut aussi contribuer à gagner en productivité grâce à la virtualisation, l'optimisation, la maîtrise énergétique et la flexibilité de l'approvisionnement.
- *Gérer les risques* : la sécurité, la résilience et la conformité sont d'ores et déjà des objectifs à atteindre dans l'environnement actuel. Les entreprises dynamiques doivent se préparer à affronter les nouveaux risques posés par un monde toujours plus connecté et collaboratif.

**Sécurité, résilience et intelligence : l'optimisation des infrastructures dynamiques au service de l'innovation stratégique**  
Conclusions des perspectives Smarter Planet

- Les DSI comprennent que pour réaliser le potentiel d'une planète plus intelligente, leurs entreprises vont devoir repenser leur infrastructure informatique au plus vite et entreprendre une modernisation à grande échelle.
- Tout type de restauration s'accompagne inévitablement de pièges ; par exemple, la tendance à investir dans ce que nous avons au lieu de ce dont nous avons besoin.
- Que peuvent faire les DSI pour prolonger la vie de l'infrastructure actuelle, tout en créant les bases d'une infrastructure plus solide, à même de soutenir l'innovation et la croissance futures ?

*Figure 1.6 : infrastructure dynamique*

## **Développement durable et au-delà**

L'environnement est une question omniprésente et urgente pour le grand public. Cette pression est renforcée par les agences gouvernementales et d'autres personnes d'influence clés qui exigent une responsabilité accrue dans les domaines qui ont un impact sur l'environnement et la société. Le développement durable doit être bénéfique autant pour le monde économique que pour l'environnement. Les entreprises privées, les organismes publics, les collectivités, les régions et des secteurs entiers de l'industrie sont confrontés à l'élaboration de stratégies et de solutions visant à devenir plus respectueux de l'environnement, d'une manière qui permette également de générer de nouvelles opportunités en termes de revenus et de réduire les coûts et les risques. La hausse des coûts énergétiques, les normes environnementales et les réglementations en vigueur poussent les DSI et les centres de données à un point de basculement, ce qui nécessite le déploiement de nouvelles stratégies et infrastructures ; le développement durable doit être le mot d'ordre pour orienter la vaste transformation de l'informatique.

L'énergie intelligente et l'amélioration de la gestion du carbone ont pour but de faire évoluer intelligemment des systèmes passifs ou « non intelligents » en vue de créer des « systèmes intelligents » considérablement plus efficaces et plus fiables, et donc à même d'économiser l'énergie et les ressources. Les réseaux de services, les systèmes de transport et les gisements de pétrole intelligents gagneront tous en efficacité. Il s'agit de solutions concrètes, disponibles aujourd'hui, qui permettent de maîtriser et d'exploiter la puissance de l'intelligence intégrée pour :

- Optimiser la gestion de l'énergie
- Limiter l'impact de notre énergie sur l'environnement et la rendre plus fiable

- Diminuer l'encombrement de la circulation et les émissions de gaz à effet de serre qui y sont associés
- Réduire la demande énergétique

**Développement durable et au-delà : faire des stratégies environnementales intelligentes un avantage concurrentiel**  
Conclusions des perspectives Smarter Planet

- Les DSI peuvent aider leurs entreprises à valoriser la démarche écologique : facteur de rentabilité et processus métier optimisés visant à réduire les coûts opérationnels, limiter l'impact environnemental de l'entreprise et accroître la productivité.
- Face au défi de juguler la consommation énergétique croissante des centres de données, les DSI doivent évaluer et étendre les investissements informatiques existants et déployer de futures infrastructures plus efficaces, dotées de contrôles renforcés, moins coûteuses et avec un moindre impact.
- Les DSI ont l'opportunité de capitaliser sur l'émergence de systèmes intelligents qui auront un impact profond sur les enjeux auxquels est confrontée la planète.
- Il est possible d'atteindre une meilleure visibilité et de réaliser des économies d'énergie à travers l'ensemble de l'entreprise en faisant évoluer intelligemment des systèmes passifs ou « non intelligents » pour créer de nouveaux systèmes plus intelligents, considérablement plus efficaces et plus fiables.

*Figure 1.7 : développement durable et au-delà*

## **Pourquoi développer une nouvelle intelligence ? Pourquoi maintenant ?**

Aujourd'hui, les entreprises ont plus d'informations à traiter que jamais, et ont pourtant du mal à tenir la cadence.

- A l'heure actuelle, le gaspillage, l'inexactitude et le nombre d'opportunités manquées représentent les problèmes majeurs d'une entreprise. Or, ils ont une cause commune : une maîtrise insuffisante des informations.
- Parallèlement à cela, presque tous les éléments de notre quotidien peuvent être traités d'un point de vue numérique et interconnectés, des routes que nous empruntons aux vêtements que nous portons, en passant par nos appareils ménagers. La technologie et le développement de réseaux abondent et sont disponibles à moindre coût, ce qui entraîne des bénéfices majeurs obtenus grâce à des dispositifs interconnectés.

- L'utilisation de davantage de dispositifs pour obtenir un plus grand volume d'informations présente une complexité accrue. Les entreprises capables de déchiffrer et de prédire les tendances émergentes gagneront en compétitivité.
- Un avantage concurrentiel est difficile à préserver lorsqu'il n'est fondé que sur des gains en termes de productivité et de rentabilité au niveau des transactions. Une automatisation simple constitue un atout, mais ne permet pas de se démarquer de la concurrence.

<b>Les défis d'une entreprise</b>		
<b>Volume</b> Chaque jour, 15 pétaoctets de nouvelles informations sont générés. On estime que le volume des informations codifiées dans le monde double actuellement toutes les 11 heures	<b>Variété</b> 80 % de la croissance de ces nouvelles données concerne un contenu non structuré, généré en grande partie par e-mail, et constitué de plus en plus de documents, images, fichiers audio et vidéo	<b>Rapidité</b> Une entreprise de 1 000 salariés dépense 5,3 millions de dollars par an pour trouver les informations stockées sur ses serveurs. 42 % des chefs d'entreprises disent utiliser de mauvaises informations au moins une fois par semaine

Figure 1.8 : les défis de la nouvelle intelligence

La nouvelle intelligence associe le potentiel des connaissances humaines aux performances de l'informatique. Elle crée un nouveau programme, basé sur la prise de conscience de la situation et *l'anticipation*. Ce concept met l'accent sur *l'optimisation*. Ainsi, il permet de passer d'une politique de sécurité publique réactive à une politique de prévention du crime proactive, d'un système de suivi et de gestion sanitaire à un processus de prévision et de prévention des épidémies, d'un procédé d'aide à la décision à une délégation de décisionnaires. Ce concept a pour but ultime l'acquisition de capacités d'anticipation, qui représentent le meilleur moyen pour aller au-delà de la notion consacrée du « sense and respond » (« sentir et réagir »). Ce monde nouveau, tel qu'il se présente à nous aujourd'hui, nous met face à une réalité : lorsque vous réagissez, il est peut-être déjà trop tard. En effet, ce monde catégorique, toujours sous pression, imprévisible et évoluant à un rythme soutenu exige que l'on *soit à la fois rapide et précis*.

La nouvelle intelligence offre aux meilleures entreprises des possibilités inédites, permettant d'optimiser les processus, de travailler en collaboration et de favoriser l'innovation.

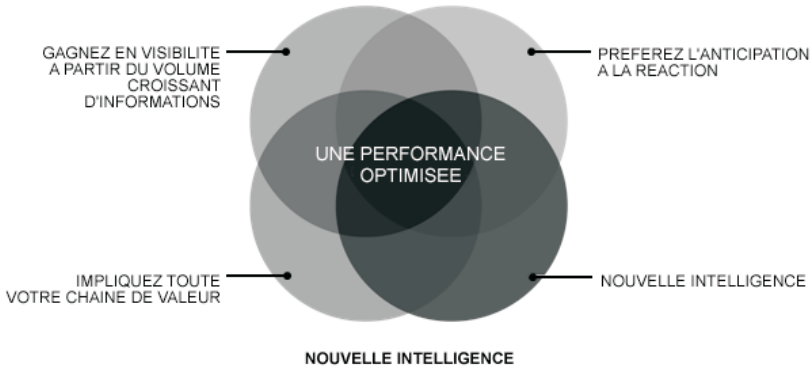


Figure 1.9 : les opportunités de la nouvelle intelligence

## Découverte de la nouvelle intelligence

Le volume des informations disponibles va toujours croissant, mais, proportionnellement, la quantité collectée, gérée, analysée et mise à disposition de ceux qui la requièrent diminue (ce phénomène étant d'autant plus conséquent pour les informations générées en temps réel). Nous franchissons une nouvelle étape, qui consiste à améliorer radicalement notre façon de saisir, traiter, présenter, évaluer, regrouper, hiérarchiser, prévoir et analyser les données émanant des plus grands systèmes économiques, sociaux et physiques du monde. Seule une solide infrastructure informatique permet d'exploiter efficacement ces nouvelles compétences dans une entreprise.

Pour faire face à ce foisonnement de données, la création de nouveaux centres de données disposant de grandes capacités de stockage et d'un traitement rapide est nécessaire, mais insuffisante en elle-même. Le flux d'informations doit être géré afin de fournir une visibilité approfondie, qui permet alors de prendre des décisions plus judicieuses, plus rapidement. Les applications de la nouvelle intelligence permettent de transformer les informations en atouts stratégiques, qui offrent la possibilité de gagner rapidement et durablement en compétitivité.

Pour faire face à une cadence au rythme toujours croissant, de plus en plus de décisions doivent être réintégrées à l'entreprise, ce qui implique plus de personnes ayant accès en *temps réel* à des informations plus précises. Or, le fait de disposer d'outils plus avancés, permettant d'explorer de manière systématique de nouvelles données structurées et non

structurées, de partager des informations et de collaborer pour la prise de décisions ne permet pas de dépasser les limites des capacités humaines. La nouvelle intelligence exigera que de plus en plus de décisions opérationnelles en temps réel soient reliées aux systèmes eux-mêmes (inventaires, flux, couvertures). Dans ce monde plus intelligent, les informations provenant des marchés, des chaînes d'approvisionnement, des clients et des meilleures applications entraînent une réaction immédiate et permettent d'agir de manière proactive.

Le passage à un système basé sur l'information peut débuter à n'importe quel stade de maîtrise des informations de l'entreprise, et offrir des avantages immédiats. Pour cela, nul besoin de réaliser des investissements considérables en termes de ressources ou de temps. Par ailleurs, vous pouvez démarrer à n'importe quel stade, en fonction de votre propre parcours. Cependant, pour bénéficier de tous les avantages d'un système basé sur l'information, vous devez considérer ces trois éléments :

- *Planifiez votre stratégie de gestion des données* : vous devez vous assurer que vous disposez d'un programme garantissant que vos informations sont en adéquation avec vos objectifs stratégiques. Vous devez notamment savoir comment appliquer au mieux ces informations à des situations caractéristiques de votre entreprise ou domaine d'activité.
- *Mettez en application vos analyses stratégiques pour optimiser vos décisions* : vous devez être en mesure de prendre des décisions plus judicieuses et plus précises, plus rapidement, et ce au moyen de planification, de suivi, de production de rapports et d'analyse de vos informations.
- *Créez une plate-forme d'informations flexible* : vous devez disposer de l'infrastructure et de la plate-forme technologique nécessaires pour répondre à vos besoins. Ceci vous garantit que les informations sont fiables, qu'elles peuvent être partagées et que toute personne les requérant peut y avoir accès de manière sécurisée, partout et à tout moment.

Que vous commenciez par planifier une stratégie, améliorer une infrastructure sous-jacente, traiter un besoin spécifique de l'entreprise ou modifier des choix tactiques, vous pouvez constater dès le commencement une amélioration du processus d'optimisation (avec un véritable retour sur investissement). En parallèle, vous pouvez poser les bases d'un ensemble de services d'information standard, qui pourra être utilisé plus facilement ultérieurement. Nous verrons ces éléments plus en détail dans les chapitres suivants.

# 2

## Nouvelle intelligence Présentation de l'analyse stratégique

L'émérgence d'une économie mondiale contraint les entreprises à se montrer plus habiles dans leurs opérations et plus innovantes dans leurs décisions. Face à l'explosion du nombre de données et à la réduction des fenêtres de traitement, les entreprises luttent pour prendre des décisions en temps réel (ou quasi en temps réel), et ainsi gagner en compétitivité.

Des informations précieuses peuvent revêtir plusieurs formes : structurées ou non structurées, opérationnelles ou transactionnelles, en temps réel ou historiques. Elles sont dispersées au sein de toute entreprise. Ces informations peuvent se trouver dans des bases et entrepôts de données, dans des e-mails et journaux de transactions, dans des journaux d'appels, dans des données relatives aux clients ou dans des demandes de réparation. Il peut également s'agir de données XML résidant dans des systèmes transactionnels, qui ne peuvent être utilisés ou analysés dans des bases de données.

Cependant, si ces données peuvent être exploitées de manière satisfaisante, les entreprises peuvent prendre de meilleures décisions pour dynamiser les ventes, améliorer les processus et les services, doper la productivité d'une équipe, réduire les risques inhérents à l'activité de l'entreprise et rationaliser les relations avec clientèle, partenaires commerciaux et fournisseurs.

L'initiative Smarter Planet permet aux entreprises de se concentrer sur la valeur, de tirer parti des opportunités et d'agir avec célérité. Mais comment atteindre ces objectifs exceptionnels ? A quelles difficultés doit-on se préparer pour élaborer un système à la



fois équipé en technologie, interconnecté et intelligent ? Nous verrons comment mettre en œuvre un tel système au chapitre 4. Nous allons maintenant examiner de plus près le cerveau d'un Smart Analytics (système d'analyse intelligent). Ceci nous permettra de comprendre le quotidien des professionnels qui utilisent ce système, et la manière dont ils tirent profit de ses formidables capacités.

Le cerveau d'un système d'analyse est, bien évidemment, l'analyse. Mais ce cerveau n'est pas confiné dans une seule partie du logiciel, un seul sous-système ou dans une base de données. Une analyse est une combinaison des éléments suivants : de nombreux processus (modélisation de prévision ou descriptive, optimisation, préparation de données) effectués au sein de l'entrepôt pour l'extraction d'informations essentielles dans un contexte de détection et de modélisation des processus métier ; l'entrepôt lui-même (plate-forme, fonctionnalités) ; la génération de rapports électroniques et les outils d'analyse (implémentation, déploiement pour les utilisateurs) ; enfin, trop rarement mentionnée, l'intelligence des personnes utilisant le système d'analyse dans le but de dégager des bénéfices pour leur activité. Ainsi, la matière grise s'associe à l'écran, à des processus analytiques et algorithmiques, ainsi qu'à l'entrepôt.

Pour la suite de ce chapitre, nous examinerons la hiérarchie des couches analytiques représentée à la figure 2.1., dans laquelle les outils destinés à l'utilisateur sont supportés par plusieurs processus analytiques. Ces processus eux-mêmes sont supportés par des systèmes IBM d'archivage et d'accélération des analyses. Cette hiérarchie est composée de plusieurs couches : interface utilisateur (outils destinés à l'utilisateur), processus analytique (processus système et algorithmes pour l'analyse) et infrastructure (entreposage et composants Smart Analytics).

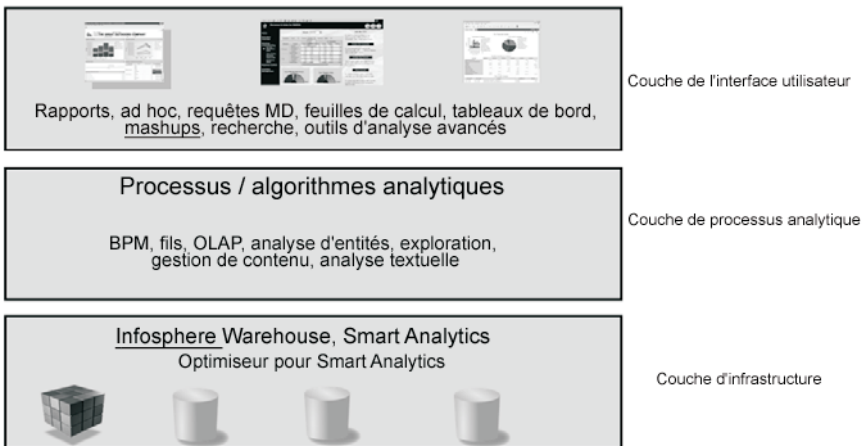


Figure 2.1 : couches analytiques

---

Un système Smart Analytics efficace est en mesure de répondre à quatre questions de l'utilisateur. Chaque réponse représente une grande valeur ajoutée et un meilleur retour sur investissement. Les outils destinés à l'utilisateur et les processus analytiques varient selon la question, tandis que l'infrastructure d'entreposage fournit une base centralisant tous les outils et analyses.

*Quelles sont les performances de mon entreprise ?* L'utilisateur consulte des rapports pré-établis et interprète les résultats. Les outils destinés à l'utilisateur prennent le plus souvent la forme de rapports pré-établis standard, mais il peut également s'agir de rapports ad hoc. Les processus analytiques sous-jacents sont essentiellement des requêtes standard ou interactives, mais ils peuvent également inclure des résultats standardisés, obtenus à partir d'analyses avancées (ex : exploration de données).

*Comment expliquer les performances de mon entreprise ?* Dans le cas présent, l'utilisateur réalise souvent ses propres calculs, allant au-delà des données issues des rapports standard ; il cherche alors à mieux comprendre les éléments à l'origine des résultats de l'entreprise. Les outils utilisés sont souvent des feuilles de calcul, des rapports ad hoc et des analyses multidimensionnelles (MD). Les processus analytiques incluent généralement des requêtes interactives et des traitements analytiques en ligne (OLAP). Ils peuvent également utiliser des résultats standardisés ou ad hoc obtenus à partir de modèles d'exploration de données préconfigurés.

*Que seraient les résultats de mon entreprise si je modifiais un paramètre spécifique ?* Ce type d'utilisateur souhaite réduire les dépenses ou augmenter le chiffre

### **Quelques précisions sur le langage XML (Extensible Markup Language)**

Ces dernières années, le langage XML s'est imposé comme un standard pour les échanges d'informations entre entreprises, ou entre les différents services ou applications d'une même société. De par sa nature flexible et extensible, le langage XML s'avère être l'outil idéal pour modéliser des informations d'entreprise semi-structurées et fréquemment modifiées. Il constitue également un standard pour les échanges d'informations via des services Web. Ses applications particulièrement utiles vont de simples processus d'échange de documents d'entreprise à des domaines hautement spécialisés, tels que les sciences de la vie.

Face à l'efficacité et à la popularité du langage XML, et au regard de ses avantages pour les analyses stratégiques et la nouvelle intelligence, IBM a pris la tête d'une industrie basée sur le support XML en mode natif. Le logiciel IBM Information Management and Analytics gère et exécute des données XML en tant que structure de niveau supérieur, prise en charge en mode natif par les moteurs et processus (et non en tant qu'extension). Le langage XML étant devenu aussi essentiel à l'analyse stratégique que les données de relations classiques, IBM a créé un support XML intrinsèque à tout élément, des mashups, processus de recherche et d'exploration de données à la gestion de contenu et aux analyses textuelles.

d'affaires, souvent sans opérer de changements au niveau opérationnel de l'entreprise. Par exemple, il peut s'agir de prévoir une augmentation des bénéfices sur la base d'une potentielle augmentation des ventes, suite à une campagne marketing. Les outils utilisés sont généralement des feuilles de calcul et des interfaces graphiques utilisateur, avec analyses multidimensionnelles et fonctionnalités d'exploration de données. Les processus analytiques peuvent associer requête, traitement analytique en ligne (OLAP) et résultats ad hoc, ces derniers étant obtenus à partir de modèles d'exploration de données. Il peut s'agir notamment d'exécution « à la volée » de processus d'exploration prédéterminés ou « guidés ».

*En quoi puis-je améliorer mon business model pour gagner en compétitivité ?* Ce scénario est le plus crucial : il s'agit de déterminer l'impact de changements au niveau comportemental et organisationnel sur la rentabilité de l'entreprise. Pour répondre à ce type de questionnement, plusieurs outils sont généralement employés. Chacun de ces outils met en œuvre des processus et algorithmes analytiques avancés au sein de l'entrepôt. Il peut s'agir par exemple d'un traitement analytique en ligne (OLAP), d'exploration de données ou d'analyses textuelles. Ces processus requièrent généralement un haut niveau de capacités d'analyses. Ils peuvent faire appel à plusieurs types de données : historiques, de prévision, relatives aux points de vente, non structurées, ou encore des données tierces (ex : démographiques ou relatives aux historiques de crédit).

## **Interactions entre utilisateur et informations**

L'utilisateur emploie de nombreux outils lui permettant de gagner en visibilité à partir du système d'analyse. Pour chaque outil ou approche, nous vous présenterons la manière de l'utiliser et les bénéfices qui en découlent en termes de visibilité au sein de l'entreprise.

### **Rapports**

Il existe de nombreux types de rapports, mais le rapport planifié est le plus courant. Il se présente souvent au format « greenbar », illustré à la figure 2.2. L'utilisateur peut demander aux services d'information à ce que des modifications soient apportées au rapport mais, dans l'ensemble, celui-ci reste le même chaque semaine. De manière générale, la création d'un tel rapport fait l'objet d'une réflexion poussée, et de nombreux processus sont exécutés par lots pour obtenir ces résultats. Ces rapports représentent un atout considérable pour l'entreprise, et l'utilisateur sait trouver directement la page correspondant aux informations souhaitées. A partir de là, il devra souvent étudier les données plus en détail ou réaliser des calculs supplémentaires, afin de mieux comprendre les résultats de l'entreprise. (Dans la plupart des entreprises, on voit fréquemment les rapports imprimés comporter des valeurs dérivées annotées.) Les rapports standard sont généralement distribués en ligne, et doivent inclure une fonction permettant à l'utilisateur

d'exporter les données vers une feuille de calcul. Sans cette fonction, l'utilisateur peut se voir contraint d'utiliser des logiciels de masquage de l'interface, à réaliser des copier-coller ou, malheureusement, à resaisir les données dans l'outil d'analyse universel qu'est la feuille de calcul.

ID produit	Nom du produit	Unités en stock	Unités commandées	Unités commandées
101	Pois congelés	147	200	150
1571	Peluches	15	10	20
1362	Horloge numérique	4	10	5
400054	Destructeurs d'odeurs	35	50	25
821	Stylo bille	101	0	75
3347	DB2 pour les Nuis	1527	1500	2000

Figure 2.2 : rapport Greenbar

L'exemple de rapport présenté à la figure 2.2. démontre qu'une partie spécifique d'un rapport peut ne pas contenir toutes les informations requises par un utilisateur ; par exemple, on ne trouve pas d'informations concernant les ventes. Ces informations peuvent être indiquées dans une autre partie du rapport, ce qui contraint l'utilisateur à le parcourir et à noter les valeurs nécessaires à la compréhension de la situation de l'entreprise.

Pour une entreprise, le grand avantage que représente un rapport standard réside dans sa capacité à être exploité par un grand nombre d'utilisateurs, et à répondre de manière très complète à de nombreuses questions. En revanche, l'utilisateur doit bien connaître le rapport pour trouver les données pertinentes. Par ailleurs, ce rapport n'aide en rien l'utilisateur à extraire les nouvelles informations. Par exemple, les Bénéfices peuvent être calculés en soustrayant le Total des dépenses du Total des ventes, mais il faut rechercher, calculer et formater les valeurs par soi-même. Les rapports standard peuvent être améliorés au moyen de données tabulaires, graphiques et tableaux attractifs d'un point de vue visuel, tels qu'illustrés à la figure 2.3.



Figure 2.3 : rapports Cognos avec graphiques et tableaux

Contrairement aux rapports greenbar, les rapports graphiques sont généralement élaborés pour un service spécifique et pour répondre à un problème donné. Il s'agit toujours de rapports prédéfinis, mais ils présentent un format bien plus favorable à la visualisation des données. Le plus souvent, ils comprennent une fonction contextuelle limitée, permettant aux graphiques de réagir lors de modifications. Les utilisateurs de ce type de rapport peuvent rapidement connaître la situation de l'entreprise et, à différents degrés, effectuer des analyses plus poussées afin de comprendre les raisons de cette situation.

## Rapports ad hoc

Lorsque l'utilisateur souhaite un rapport bien spécifique, généralement pour un usage ponctuel, il peut soit générer le nouveau rapport à partir d'un rapport standard, soit émettre sa propre requête pour les seules données nécessaires, à l'aide d'un rapport ad hoc. Nous apprécions tous de pouvoir comprendre les choses en profondeur ; or « ad hoc » est un terme latin signifiant « dans ce but ». L'utilisateur collecte souvent plusieurs modèles de rapports ad hoc, et modifie simplement quelques paramètres pour créer un nouveau rapport.

Lorsqu'un utilisateur cherche à déterminer la situation d'une entreprise, il a donc plusieurs moyens d'obtenir des réponses à ses questions. Les données peuvent provenir d'un rapport standard (directement ou à partir de calculs), ou être produites au moyen

d'un rapport ad hoc, nouvellement créé ou modifié. Il est impossible d'affirmer catégoriquement quelle méthode est la meilleure, la réponse dépendant de la complexité du calcul et de la possibilité de calculer le résultat à partir de rapports standard et ad hoc.

La plupart des outils graphiques pour les analyses ad hoc sont faciles à utiliser et plutôt intuitifs. Ainsi, l'utilisateur peut créer des snapshots rapides des parties de l'entreprise qu'il souhaite examiner. Les résultats obtenus à partir des outils de Cognos sont spectaculaires, offrant des fonctionnalités de formatage et de création de tableaux complètes, telles qu'illustrées à la figure 2.4.

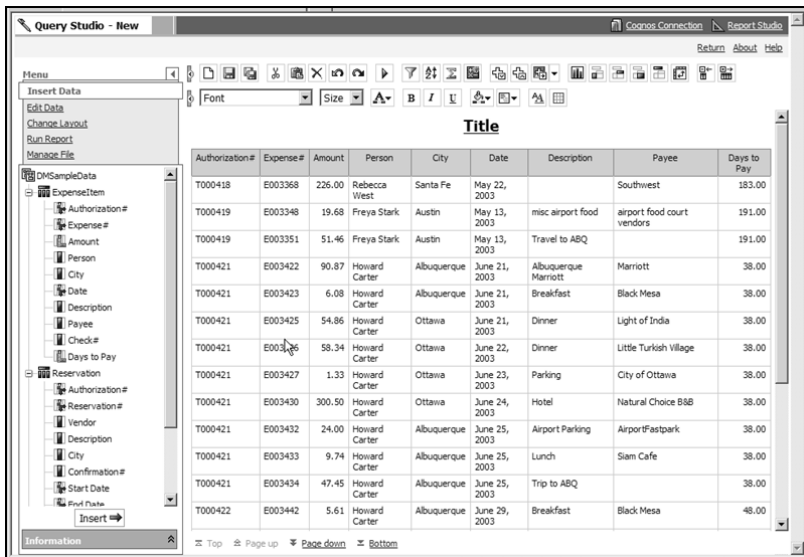


Figure 2.4 : outils de Cognos pour l'analyse ad hoc

Lorsqu'ils sont associés, les rapports standard et les analyses ad hoc représentent une grande capacité d'analyse, nécessaire à certains utilisateurs. Cependant, si cette capacité peut répondre aux questions telles que « Quelles sont les performances de mon entreprise ? », elle ne peut apporter de réponse aussi simplement à « Comment expliquer les performances de mon entreprise ? ». L'utilisateur doit réaliser une analyse plus avancée des données de l'entreprise, afin de trouver les informations qui lui permettront de répondre à la dernière question. Les rapports standard et ad hoc ne permettent pas en eux-mêmes de répondre à ces questions. Les utilisateurs ont donc pris l'habitude de se servir de feuilles de calcul, souvent remplies de données copiées à partir de rapports et de requêtes ad hoc.

## Feuilles de calcul

Un grand nombre des questions portant sur « comment et pourquoi » peuvent trouver leurs réponses à l'aide de macros, de tableaux croisés dynamiques et de tableaux, utilisés dans des feuilles de calcul. Comme cela a été indiqué plus haut, les utilisateurs ont développé de nombreuses méthodes pour charger des feuilles de calcul, et une partie a acquis des compétences certaines en analyse de données dans des feuilles de calcul. Le problème se pose à présent en termes d'interconnexion et d'intelligence. En effet, les données de l'entrepôt et les feuilles de calcul ne sont pas connectées, et la seule intelligence est celle contenue dans les macros des feuilles de calcul.

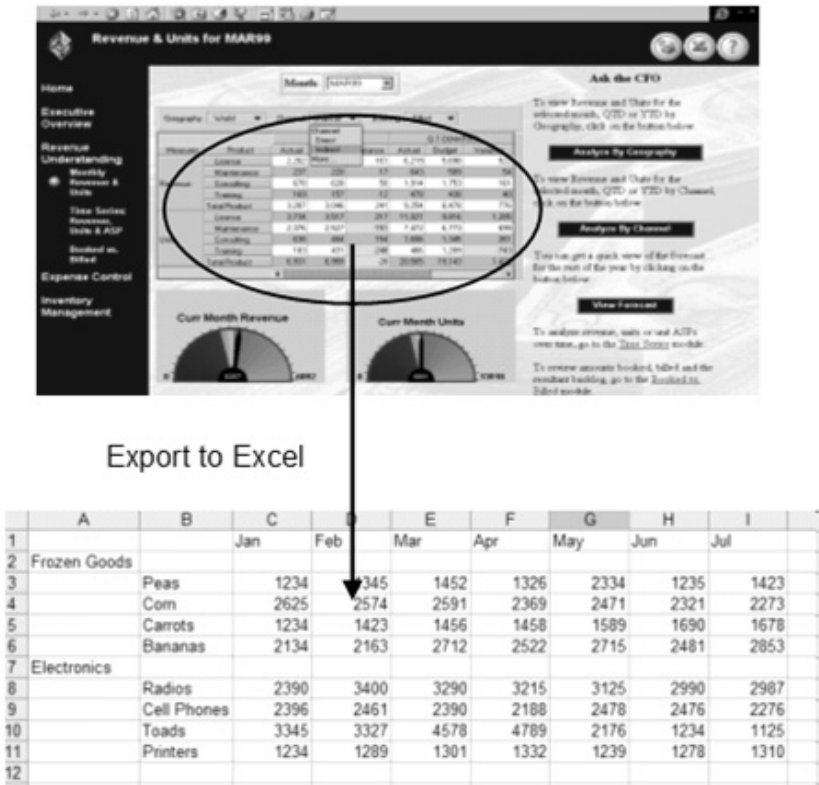


Figure 2.5 : chargement d'Excel à partir de rapports standard

De prime abord, on pourrait se demander pourquoi un utilisateur transférerait des données d'un rapport graphique à une feuille de calcul. Et si l'utilisateur en restait là, tout le monde serait d'accord pour affirmer qu'il s'agit d'une mauvaise idée. Depuis l'introduction de la feuille de calcul électronique, les utilisateurs associent des données

issues de sources multiples dans un seul document, afin d'obtenir une nouvelle visibilité de leur entreprise, en utilisant leurs propres calculs.

## **Analyse multidimensionnelle**

L'analyse multidimensionnelle est une méthode courante permettant de déterminer le comportement de l'entreprise. Toutefois, ce terme revêt plusieurs significations, et ses origines remontent aux débuts de la programmation. A la fin des années 1960, IBM a lancé le langage de programmation APL, doté d'un ensemble complet de variables et d'opérateurs multidimensionnels complexes mais efficaces. APL est passé pratiquement inaperçu, mais il a refait surface avec les analyses multidimensionnelles, lorsque la feuille de calcul est devenue populaire. Mais qu'est-ce qu'une analyse multidimensionnelle ?

L'analyse multidimensionnelle est devenue le moyen de calculer une mesure de l'entreprise (ex : ventes, marge, bénéfices, dépenses) par rapport à de nombreuses autres dimensions (telles que l'heure, le produit et l'emplacement). L'expression résumant ce procédé ressemble généralement à : « Quels sont mes bénéfices par heure, consommateur, produit, canal de distribution ? » Chaque critère ajoute une dimension supplémentaire de réflexion. Ainsi, l'exemple ci-dessus est une requête à quatre dimensions : mesure de l'entreprise, heure, produit et canal de distribution.

Chaque dimension possède une hiérarchie naturelle, conçue pour refléter le fonctionnement de l'entreprise, ainsi que la manière dont l'analyste perçoit cette dimension. Le meilleur moyen de comprendre une hiérarchie dimensionnelle est de regarder la notion de temps, dont la hiérarchie peut inclure Année, Trimestre, Mois, Semaine et Jour. L'analyste peut observer le niveau Mois, mais remontera ou descendra facilement vers le niveau Trimestre ou Semaine. Dans l'exemple suivant, l'analyste a défini cinq dimensions sous Excel. La dimension Produit est imbriquée dans la dimension Mesures de l'entreprise, et les dimensions Heure et Données réelles/Budget sont imbriquées. Chaque dimension apparaît à un ou plusieurs niveaux de sa hiérarchie. Les produits se trouvent à deux niveaux.



	A	B	C	D	E	F	G	H	I	J
1						Région 1				
2			Janv.		Fév.		Mars.		1er trim	
3			Données réelles	Budget	Données réelles	Budget	Données réelles	Budget	Données réelles	Budget
4	Débardeurs	Ventes	1812	1690	1754	1640	1805	1690	5371	5020
5		Coûts	599	550	588	540	596	540	1783	1630
6		Marge	1213	1140	1166	1100	1209	1150	3588	3390
7		Bénéfice	837	860	792	820	832	870	2461	2550
8	Pulls	Ventes	200	190	206	190	214	200	620	580
9		Coûts	84	80	86	80	89	80	259	240
10		Marge	116	110	120	110	125	120	361	340
11		Bénéfice	67	70	71	70	74	80	212	220
12	Vestes	Ventes	93	80	101	90	107	100	301	270
13		Coûts	38	30	41	30	43	40	122	100
14		Marge	55	50	60	60	64	60	179	170
15		Bénéfice	20	30	25	40	29	40	74	110
16	Sweatshirts	Ventes	2105	1960	2061	1920	2126	1990	6292	5870
17		Coûts	721	660	715	650	728	660	2164	1970
18		Marge	1384	1300	1346	1270	1398	1330	4128	3900
19		Bénéfice	924	960	888	930	935	990	2747	2880

Figure 2.6 : exemple Excel d'analyse multidimensionnelle

L'analyse multidimensionnelle est une méthode puissante. Elle permet en effet à l'analyste de consulter les dimensions selon divers niveaux de précision, et de croiser les dimensions afin de faire parler les données.

Depuis l'apparition de Visicalc au début des années 1980, les consommateurs de données s'appuient volontiers sur la feuille de calcul comme outil d'analyse privilégié, le format le plus fréquent étant multidimensionnel.

Les analystes se sont tournés vers la feuille de calcul, confrontés aux lacunes des méthodes de reporting classiques qui ne leur permettaient pas d'obtenir les données dont ils avaient besoin. Leurs activités nécessitaient d'analyser les données et de consulter des hiérarchies selon divers niveaux de précision et selon leurs propres critères. Pour cela, ils effectuaient à l'époque leurs propres extractions, transformations et chargements à partir de plusieurs sources (généralement à partir des rapports dont ils disposaient), en ajoutant les données de bases de données de transactions et d'entrepôts de données. Ils réalisaient les calculs analytiques directement dans la feuille de calcul à l'aide de macros et d'agrégations de feuilles de calcul. Chaque feuille de calcul équivaut à un petit magasin de données, avec ses outils ETL (extraction, conversion et chargement de données) et d'analyse embarquée.

Le nombre de feuilles de calcul suivies par les services des finances s'élève couramment à une dizaine de milliers. D'un point de vue global, cette situation est désastreuse : ce sont des milliers, voire des dizaines de milliers de magasins de données, de processus ETL non documentés et de calculs analytiques propres à chaque analyste. Pas étonnant

que les analystes se plaignent de l'incohérence des chiffres, et que chacun croit posséder les chiffres exacts !

Comme le montre la figure 2.7, l'ajout de graphiques et de tableaux permet d'obtenir des outils d'analyse multidimensionnelle relativement esthétiques. Cognos permet de découper les données en tranches (slicing) ou en dés (dicing) à partir de tableaux synchronisés.

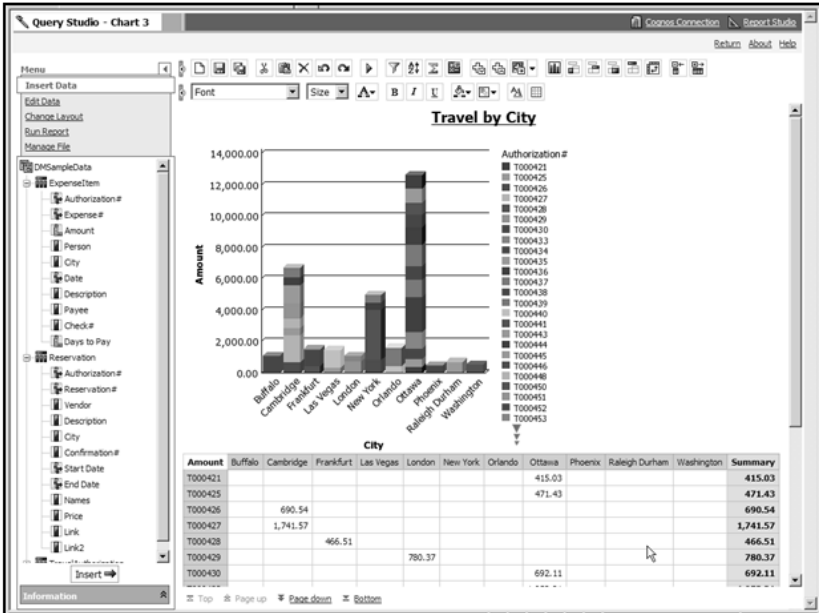


Figure 2.7 : outils graphiques Cognos pour l'analyse multidimensionnelle

Grâce à l'émergence et à la maturation rapides des technologies basées sur le Web et les services, les environnements d'analyse bénéficient de fonctions essentielles qui permettent de satisfaire les exigences croissantes en « nouvelle intelligence ». Diverses technologies ultra dynamiques, adaptatives, personnalisées et de réseaux sociaux (tableaux de bord, mashups avec fonctions Web 2.0, moteurs de recherche d'entreprise, etc.), s'associent pour fournir aux décideurs, plus interconnectés que jamais, la richesse d'information nécessaire pour accélérer la prise de décisions avisées. Voici une vue d'ensemble de ces outils graphiques nouvelle génération.

## Tableaux de bord

Les tableaux de bord, qu'ils soient déployés à l'échelle de l'entreprise ou personnalisés pour un groupe, voire pour un seul utilisateur, constituent aujourd'hui un point d'information standard de l'industrie à partir duquel la quasi-totalité de l'activité

d'analyse est amorcée. Le tableau de bord est la « page d'accueil » essentielle, la première chose que l'utilisateur consulte en début de journée, puis régulièrement au fil des heures. Grâce à la possibilité d'envoyer des informations sur presque n'importe quel type de périphérique mobile ou de canal, les tableaux de bord sont disponibles et accessibles 24 h/24, 7 j/7, à tout moment et en tous lieux.

Le tableau de bord analytique, tout comme son proche cousin la *scorecard*, offre un aperçu synthétique, interactif et densément graphique sur un seul écran, des indicateurs de performances clés et des informations stratégiques en temps quasi-réel. Intégré dans l'environnement d'analyse, le tableau de bord rassemble des informations essentielles à la fois dans le contexte de l'utilisateur qui consomme l'information, et dans le contexte du processus métier exécuté par l'utilisateur. Le tableau de bord n'est pas un simple vecteur d'information. Il permet à l'utilisateur d'agir par le biais de la navigation et de liens interactifs. La figure 2.8 présente un exemple de tableau de bord.

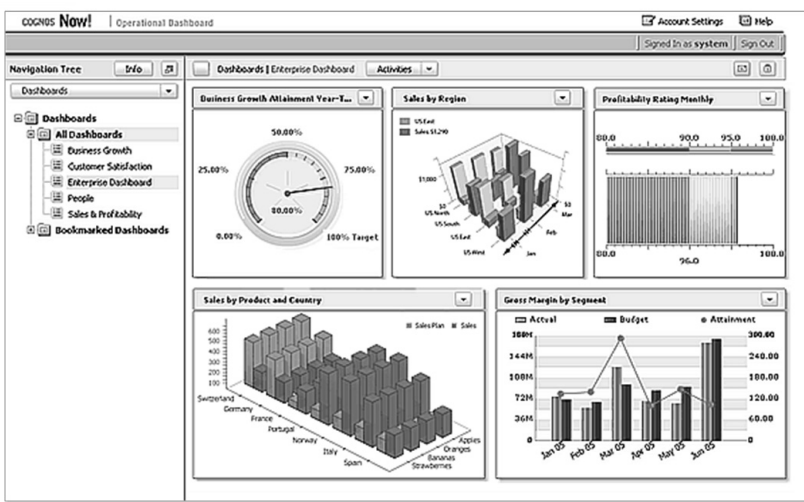


Figure 2.8 : exemple de tableau de bord

Bien qu'ils puissent varier considérablement les uns par rapport aux autres, les tableaux de bord partagent des caractéristiques communes :

- *Plusieurs sources d'informations en contexte* : pour être utiles, les tableaux de bord rassemblent diverses informations et données d'analyse issues de plusieurs sources dans une même vue synthétique, à la fois dans le contexte des besoins qu'elles viennent satisfaire, et les uns par rapport aux autres. Ces sources peuvent être internes et externes à l'entreprise.

- *Accès permanent aux données d'analyse* : les tableaux de bord permettent d'interconnecter l'utilisateur et l'environnement d'analyse par le biais de points d'accès à toute une gamme de ressources et d'applications. Le tableau de bord donne accès à de nombreux éléments (rapports prêts à l'emploi, réseau de cubes OLAP, fonctions avancées de prévision et d'exploration de données, etc.) de manière et dans un contexte facilement consommables par l'utilisateur. Les tableaux de bord constituent un environnement utilisateur cohérent, quels que soient le contenu et les processus complexes qui le sous-tendent. Principalement basés sur le Web, les tableaux de bord proposent à l'utilisateur final des informations d'analyse sur un outil privilégié : le navigateur Web.
- *Reliés et interactifs* : pour dépasser l'information passive, les tableaux de bord doivent être reliés et extrêmement interactifs. Des contrôleurs et indicateurs de performances sont reliés dynamiquement pour donner accès à l'utilisateur, en un simple clic, à des rapports plus détaillés, des applications ou des processus métier. Le tableau de bord relie entre elles les représentations de données : en cliquant sur un détail d'un graphique (par exemple, une catégorie de produit d'un graphique circulaire de vente), l'utilisateur actualise simultanément le contenu d'autres graphiques (par exemple, la liste des principaux fournisseurs de produits d'un autre graphique).
- *Communication d'informations de tous types, à tous et en tous lieux* : ultra personnalisables, les tableaux de bord permettent d'afficher les informations et données d'analyse les plus utiles aux utilisateurs et aux différentes communautés. Les tableaux de bord peuvent également être configurés pour envoyer du contenu et des alertes vers presque n'importe quel type de support et de périphérique sur réseau local ou distant.

Les tableaux de bord constituent *le* portail le plus plébiscité pour le lancement et l'exécution de la plupart des activités d'analyse et décisionnelles. En rassemblant en toute transparence des informations et des applications disparates d'information et de business intelligence (BI) sur un point focal unique, les tableaux de bord permettent aux professionnels d'accélérer la prise de décisions avisées.

## ***Mashups et Web 2.0***

L'apparition récente des mashups et des technologies Web 2.0 associées a dynamisé les solutions de tableaux de bord classiques, grâce à l'interconnectivité et aux fonctions de collaboration d'outils de réseaux sociaux très en vogue, et à la possibilité pour chaque utilisateur de réaliser un développement et un déploiement rapides.

En apparence, les pages de mashup ressemblent beaucoup à un tableau de bord classique. Mais les mashups sont en réalité très spécifiques et extrêmement bien adaptés au monde de la « nouvelle intelligence » grâce à leur technologie. Basés sur la technologie Web 2.0 nouvelle génération, les mashups sont compatibles avec les interfaces individuelles personnalisées et offrent un excellent niveau d'information et d'interconnectivité utilisateur, qui font évoluer les tableaux de bord. Les pages de mashup sont assemblées par l'utilisateur à partir de sources de données et de widgets développés, gérés et personnalisés par la communauté d'utilisateurs selon des standards reconnus et des processus légers. Dans cet environnement, chacun a la possibilité de développer individuellement, ou en collaboration avec des dizaines, voire des centaines d'autres utilisateurs, des tableaux de bord hautement personnalisés adaptés à l'individu, sans posséder de compétences approfondies du développement logiciel ni subir de cycles de développement interminables.

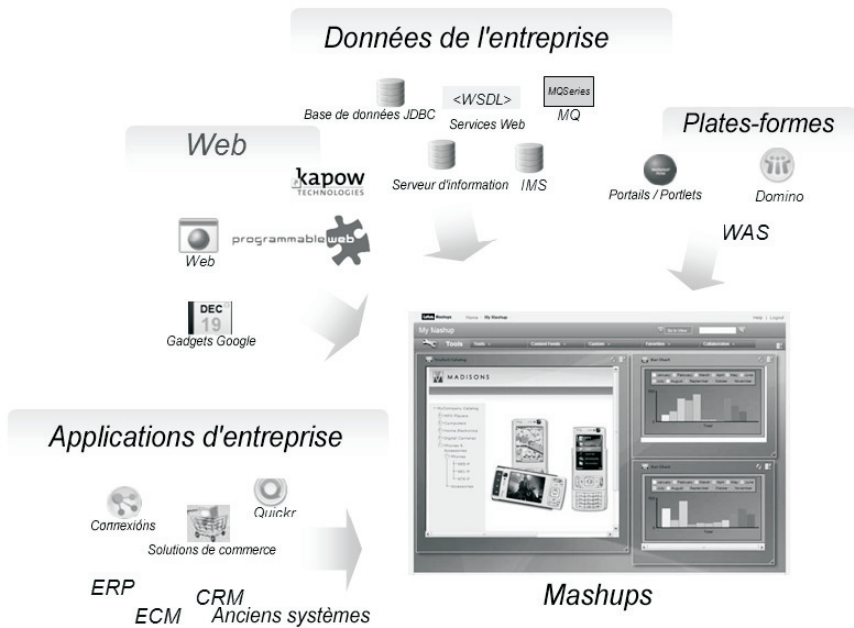


Figure 2.9 : les mashups sont alimentés à partir de diverses sources de données

Comme le montre la figure 2.9, des sources de données peuvent être créées ou adaptées pour alimenter n'importe quel type de données, d'informations ou d'analyses, d'où qu'elles proviennent, notamment des sources internes et externes. Exemples : bases de données relationnelles, cubes OLAP, applications de services Web, sources XML, feuilles de calcul, sources d'actualités commerciales, services publics, etc. Les sources de

données peuvent être associées, transformées ou enrichies selon les besoins pour synthétiser les informations nécessaires. Les données de tous types provenant de toutes les sources sont dorénavant interconnectées ! De même, des widgets graphiques et textuels peuvent être créés ou adaptés à l'aide d'outils relativement simples, afin de présenter les informations dans le format souhaité, ou pour activer la visualisation graphique interactive. Comme les tableaux de bord, les pages de mashup peuvent être interactives, reliées dynamiquement et contextuelles. Contrairement aux tableaux de bord classiques, les mashups offrent à l'utilisateur un moyen pratique de créer rapidement une visualisation personnelle des informations dont il a besoin. Ceci permet à chaque individu de bénéficier d'un capital intelligence inégalé, disponible à tous dans l'entreprise.

Descendants du phénomène des réseaux sociaux, les mashups se prêtent naturellement à la collaboration et à l'interconnectivité humaine. Les mashups permettent de générer des annotations et des évaluations, et comportent des référentiels et des catalogues de sources de données et de widgets. Les mashups sont une véritable évolution pour les tableaux de bord, et répondent directement aux défis de la « nouvelle intelligence » en termes de diversité et de rapidité.

## **Recherche**

Dans le monde interconnecté de la « nouvelle intelligence », les technologies de recherche évoluées ont ouvert de nouveaux horizons en matière de diversité, de rapidité et de volume, et permettent à davantage d'utilisateurs de prendre des décisions plus avisées. A leurs débuts, les moteurs de recherche parcouraient le World Wide Web en quête de pages à indexer. Bien que toujours dédiés à la recherche de documents portant sur des sujets divers, les processus de recherche visent aujourd'hui des informations utiles à l'utilisateur dans n'importe quel contexte.

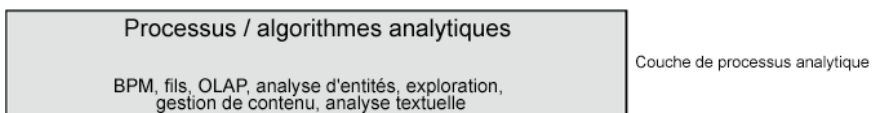
Bien qu'il soit toujours dédié à la recherche de documents sur des sujets spécifiques, le processus de recherche vise également aujourd'hui la recherche d'informations pertinentes pour l'utilisateur dans le contexte de processus métier. Les sources et formats d'information se sont considérablement diversifiés, bien au-delà des tables, colonnes et indicateurs bien structurés des environnements d'analyse classiques. Selon des études fréquemment citées, environ 80 % des nouveaux contenus de données ne sont pas structurés, c'est-à-dire qu'ils se trouvent dans des documents, e-mails, notes, journaux d'appels, articles de presse, blogs, communications instantanées, images, vidéos, journaux d'applications et données non structurées (XML, par exemple) de toutes sortes. Cette masse considérable de contenu renferme de précieuses informations utiles et essentielles à la prise de décisions avisées, et doit être exploitée. Associé aux indicateurs structurés classiques, ce contenu permet de faire parler les données rapidement et avec une précision optimale, afin de tirer des conclusions utiles.

La recherche est donc aujourd'hui étroitement intégrée à la distribution permanente de contenu et d'applications d'analyse, la plupart du temps par le biais de tableaux de bord et de mashups. Tandis que l'utilisateur dispose de sources d'information variées rassemblées et présentées intelligemment sur une même page, la recherche permet de trouver davantage de contenu au coup par coup. Par exemple, après avoir cliqué sur le symbole boursier d'un concurrent, l'utilisateur pourra trouver les toutes dernières informations sur les annonces de produits ou les acquisitions de ce concurrent. Ou bien en consultant en détail un tableau de stock problématique, l'utilisateur recherchera les journaux d'expédition, qui révéleront un schéma récurrent de marchandises endommagées. Enfin, une annonce de produit concurrent pourra mener à la recherche de brevets technologiques associés. Face aux tableaux de bord et aux mashups qui proposent des liens vers du contenu et des applications, la recherche se distingue par la nature imprévisible des requêtes. La recherche permet à l'utilisateur d'aller là où il a besoin, et non là où les concepteurs d'applications le veulent.

La nature adaptative et extensible de la recherche se prête particulièrement à la « nouvelle intelligence ». La recherche « apprend » à partir du schéma de comportement de l'utilisateur et de sa façon de naviguer, et personnalise les résultats de manière proactive. L'utilisateur peut ainsi accéder plus rapidement aux informations les plus utiles et qui répondent le mieux à ses besoins. Les chapitres suivants (qui présentent le contenu d'entreprise et l'analyse textuelle) décrivent la technologie qui sous-tend la recherche intelligente. La puissance de recherche des environnements d'analyse s'appuie sur le développement des capacités d'analyse textuelle.

## **Infrastructure d'analyse dédiée à la nouvelle intelligence**

La communication de données analytiques à l'utilisateur par le biais de l'environnement d'analyse repose sur la puissance de l'infrastructure d'analyse, c'est-à-dire sur les moteurs qui accèdent aux données, les transforment, les enrichissent, les consolident, les agrègent et les synthétisent sous forme d'informations exploitables, quelle que soit leur source. Ce chapitre présente les processus d'activation de l'analyse (figure 2.10) sur lesquels reposent les activités quotidiennes des analystes.



*Figure 2.10 : couche de processus analytique*

## Gestion des processus métier (BPM)

La gestion des processus métier (BPM) associe fonctions logicielles et compétences métier par le biais des personnes, des systèmes et de l'information, afin de réduire le délai entre les améliorations des processus pour favoriser l'innovation dans l'entreprise. L'intégration de la BI directement dans les processus métier optimise l'efficacité des points décisionnels humains. Lorsqu'un flux de travail est acheminé vers un décideur, il peut être accompagné d'informations pertinentes et à jour sur lesquelles le processus s'appuie. L'élimination de l'étape de recherche d'informations ad hoc permet aux décideurs d'agir rapidement et d'accélérer la prise de décisions.

L'intégration de la BI dans les processus métier contribue également à la qualité et la cohérence des décisions. En basant la prise de décisions sur des informations cohérentes et à jour, les entreprises limitent la variabilité des décisions humaines, quels que soient les décideurs. Grâce à l'intégration directe de l'information dans le flux de travail, les entreprises s'assurent que l'information sera toujours consultée avant toute prise de décision. La gestion des processus métier (BPM) dépasse la sphère logicielle et intègre les compétences et les ressources de processus (voir figure 2.11).

Composants de la gestion des processus métier (BPM)

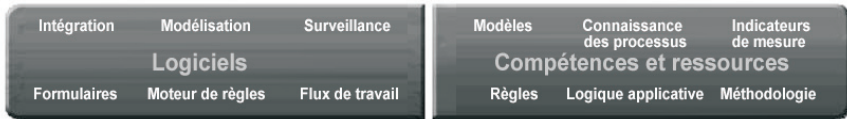


Figure 2.11 : la gestion des processus métier régit les activités opérationnelles et organisationnelles

Pour mettre en œuvre des processus de prise de décision améliorés et enrichis grâce à la BI, les entreprises doivent déterminer soigneusement quelles informations sont les plus importantes et les plus pertinentes pour la prise de chaque décision. Les analystes chargés de créer et de modéliser les processus doivent exploiter l'information à l'échelle de l'entreprise de sorte à faire le lien entre indicateurs de mesure décisionnels et résultats commerciaux. Par exemple, dans le cadre d'une décision de vente croisée/vente incitative, l'analyste doit voir au-delà des objectifs commerciaux immédiats, afin d'étudier les implications de la valeur économique du client. Trouver la bonne information peut représenter un véritable défi. C'est pourquoi les entreprises doivent encourager la collaboration entre propriétaires de processus et professionnels de la gestion de l'information pour optimiser les résultats. L'information doit également être intégrée dans le cycle d'amélioration des processus. La valeur de l'information qui sous-tend la prise de décision doit être en permanence réévaluée et réajustée à mesure de l'analyse, de l'amélioration et du redéploiement des processus.



Le meilleur moyen de doper l'efficacité des processus métier impliquant des points décisionnels complexes consiste à automatiser les décisions. Cependant, étant donnée la complexité de certaines décisions dans l'entreprise, l'automatisation n'est pas toujours une solution viable. L'utilisation à la fois d'indicateurs de mesure de processus et de BI peut toutefois permettre d'automatiser certaines parties de décisions.

L'automatisation décisionnelle repose grandement sur les données de processus d'une part, et d'autre part, sur les données de l'entreprise dans son ensemble. Les analystes doivent en effet intégrer les connaissances acquises à partir de décisions antérieures, tout en prenant en compte l'impact sur les résultats de l'entreprise. Exemple : dans une banque, les données de processus montrent que les responsables du crédit ont accordé des prêts dans 80 % des cas aux candidats ayant un score d'éligibilité supérieur à 600. Le rapprochement de ce pourcentage avec les résultats de la banque montre que le remboursement de 8 % de ces prêts s'est avéré défaillant, mais que les clients associés à d'autres produits financiers de la banque étaient beaucoup moins susceptibles de faillir aux remboursements.

En rapprochant les indicateurs clés issus des processus métier et du système de BI, il est possible de créer des règles d'automatisation de certaines parties du flux de travail. Poursuivons l'exemple des prêts : l'analyste commercial peut créer une règle de validation automatique des demandes de prêt lorsque le score du candidat est supérieur à 600, et lorsqu'il utilise un ou plusieurs autres services de la banque. Les autres demandes seront toujours soumises à la décision des responsables du crédit.

Les entreprises peuvent améliorer considérablement l'efficacité des processus en automatisant ne serait-ce qu'une partie de la prise de décisions complexes. En automatisant certaines décisions opérationnelles, les employés peuvent procéder à une gestion par exception, et se concentrer sur des activités davantage critiques et stratégiques. Ce type d'amélioration des processus n'est pas toujours réalisable. Mais en s'appuyant sur des historiques décisionnels et des données d'entreprise, il est possible de mettre au jour des indicateurs qui permettront de guider l'automatisation des flux de travail en passant par différents points décisionnels.

La surveillance des activités stratégiques (BAM) est une composante essentielle des déploiements BPM. Elle fournit une visibilité à l'intérieur des processus en vol et des applications d'entreprise. Sans surveillance, l'automatisation des processus métier présente des risques, et l'optimisation est impossible. Une vue en temps réel sur les performances des processus est primordiale pour fonctionner correctement. Cette vue peut être enrichie d'informations et d'observations supplémentaires issues de la BI.

L'instauration d'un lien entre la BAM et la BI permet aux utilisateurs de consulter rapidement les indicateurs de processus, et d'analyser les impacts à plus grande échelle des performances de processus. Cette analyse est extrêmement utile pour décider des

mesures correctives à prendre. Exemple : un indicateur clé de performances (KPI) affiché sur le tableau de bord d'une société d'assurance indique que le volume de traitement des déclarations de sinistre est inférieur aux objectifs de niveaux de service. Pour savoir quelle mesure corrective doit être prise, l'analyste peut consulter en détail les données de processus pour accéder au délai et au volume de traitement par type de sinistre. En même temps, d'autres données d'entreprise fournies par le biais de la BI peuvent révéler l'impact du volume de traitement sur la satisfaction des clients, ce qui permet de prévoir l'impact éventuel des performances du processus sur les résultats financiers. Ce contexte supplémentaire permet aux analystes commerciaux d'évaluer avec précision les coûts et avantages des réactions aux alertes de KPI. La BI peut également être utile pour déterminer les KPI à surveiller. En rapprochant les performances de processus et les résultats de l'entreprise, les analystes commerciaux peuvent identifier les KPI dont l'impact sur les résultats est le plus fort.

La BI permet aux entreprises d'enrichir et d'améliorer de nombreuses manières les processus métier et les initiatives de gestion des processus métier (BPM). Il existe divers moyens d'étendre la BPM et les processus métier aux interfaces de BI, et donc à ses utilisateurs, pour optimiser la réactivité organisationnelle.

Les données de transactions et d'historique sont une précieuse source potentielle de visibilité sur l'entreprise, dont la BI permet de consulter en détail et d'analyser les données. Les analystes peuvent ainsi mettre au jour des détails extrêmement utiles qui détermineront les priorités et règles de l'entreprise. Pour bénéficier de cette visibilité, les entreprises doivent agir immédiatement. Mais elles ne disposent généralement pas des liens directs nécessaires entre opérations et analyses. Habituellement, les analystes communiquent leurs résultats par le biais de canaux ad hoc, puis les responsables doivent prendre des décisions et mettre en place des mesures sur la base d'informations incomplètes.

Par la simple extension des données de processus au système de BI, la BPM devient immédiatement profitable aux utilisateurs de la BI. Les données d'entreprise et de transactions traditionnellement analysées à partir de systèmes de BI proviennent essentiellement de processus métier, dont les données constituent un contexte utile pour l'analyse des données d'entreprise et de transactions. Exemple : l'analyse des données d'entreprise d'un grossiste révèle l'accroissement de l'indicateur « nombre de jours d'impayé ». Cet indicateur financier provient directement des processus de commande-encaissement. En rapprochant les données de processus et les résultats financiers, l'analyste peut consulter en détail les éléments du processus pour déterminer leur impact sur les résultats.

L'objectif essentiel d'une mise en œuvre de la BI consiste à exploiter la visibilité pour améliorer les performances. Cependant, les systèmes de BI ne sont pas toujours en mesure d'accéder aux données de processus clés en temps réel révélant les rouages de la

performance. Grâce à Smart Analytics, l'extension des données de BPM au système de BI permet aux entreprises de prendre de meilleures décisions stratégiques basées sur un aperçu complet et cohérent de l'entreprise.

## **IBM InfoSphere Streams**

IBM InfoSphere Streams fournit des technologies de pointe permettant de générer et de gérer de manière extrêmement efficace des informations et des connaissances à partir de données ad hoc devant être extraites de volumes gigantesques de données potentiellement sans intérêt. InfoSphere Streams vise précisément à étendre radicalement les capacités de pointe du traitement de l'information, tout en répondant à divers défis techniques, notamment :

- Réaction rapide aux événements et aux besoins changeants
- Analyse continue des données à des débits supérieurs à ceux des systèmes existants
- Adaptation rapide aux nouveaux formats et types de données
- Gestion de la haute disponibilité, de l'hétérogénéité et de la répartition dans le cadre du nouveau paradigme de flux
- Sécurité et confidentialité des données partagées

*Le calcul des flux* constitue un nouveau paradigme. Dans le traitement « classique », des requêtes sont exécutées sur des données relativement statiques : on peut par exemple « répertorier le personnel qui habite dans un rayon de 50 km de Paris ». Cette requête renverra un seul ensemble de résultats. Avec le calcul des flux, il est possible d'exécuter un processus semblable à une « requête continue » pour identifier le personnel habitant dans un rayon de 50 km de Paris, tout en obtenant des résultats à jour et continus, les données de localisation GPS étant régulièrement actualisées. Dans le premier cas, des données statiques sont interrogées, et dans le second cas, les données sont évaluées en permanence par des requêtes statiques. InfoSphere Streams offre une évolution supplémentaire en permettant de modifier les requêtes continues dans le temps. La figure 2.12 décrit la différence entre les deux types de requêtes.

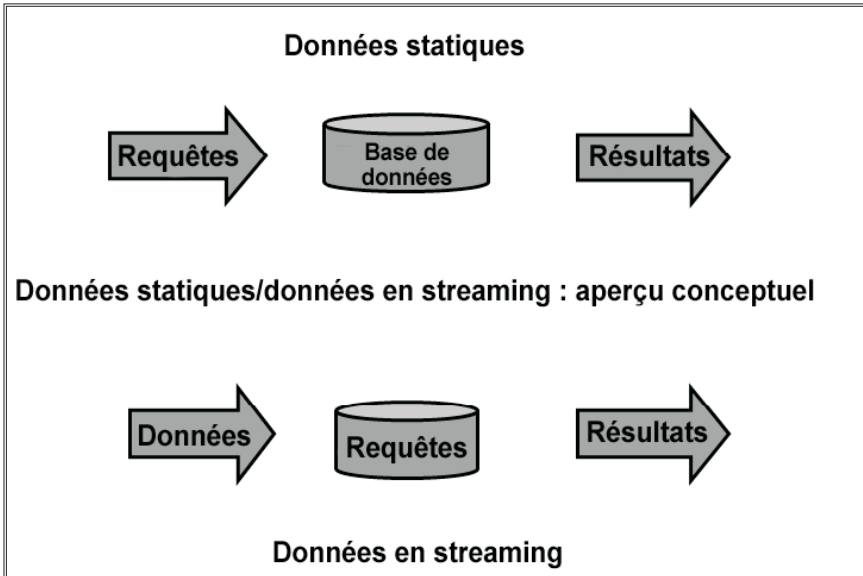


Figure 2.12 : différence entre traitement des flux et traitement classique

Contrairement à certains systèmes qui exploitent le paradigme de calcul des flux, InfoSphere Streams utilise une approche fondamentalement différente du traitement continu, en se différenciant de sa plate-forme d'exécution distribuée, de son modèle de programmation et de ses outils de développement d'applications de traitement continu. Les flux de données consommables par InfoSphere Streams peuvent provenir de capteurs, caméras, flux d'actualités, afficheurs boursiers ou de diverses autres sources, y compris de bases de données classiques.

### Cas d'usage émergent

A mesure de la généralisation d'InfoSphere Streams, un certain nombre de nouvelles applications sont explorées. Voici une synthèse des expérimentations menées par IBM mettant en avant les types d'usages qu'InfoSphere Streams est capable de mettre en œuvre.

- **Radioastronomie** : un des points forts essentiels d'InfoSphere Streams est sa capacité à exécuter des analyses sur des flux de gros volumes de données pour identifier les quelques éléments qui méritent d'être étudiés plus avant. La radioastronomie fait partie des disciplines concernées. Un certain nombre de projets reçoivent des flux de télémétrie continus provenant de radiotélescopes. Les radiotélescopes peuvent déployer des milliers, voire des dizaines de milliers d'antennes, chacune acheminant des flux de données vers un superordinateur central pour l'étude d'un point de l'univers. La solution d'infrastructure InfoSphere Streams exécutée sur ce superordinateur propose une approche plus

souple du traitement de ces flux de données. IBM travaille en collaboration avec une université pour développer des systèmes d'analyse capables d'identifier les comportements anormaux ou transitoires, par exemple les jaillissements de rayons cosmiques à haute énergie.

- *Négoce d'énergie* : l'expérimentation menée dans le domaine du négoce d'énergie montre les capacités d'InfoSphere Streams dans ce secteur. Le système expérimenté fournit aux négociants en énergie une analyse en temps réel, et indique la corrélation des événements qui impactent les marchés de l'énergie, pour permettre la prise de décisions avisées plus rapidement qu'auparavant. L'analyse comprend divers éléments : cartes de chaleur, modèles de demande énergétique, analyse technique des contrats à terme sur produits énergétiques (bande de Bollinger, prix moyen pondéré en fonction du volume, etc.), analyse de flux d'actualités pour identifier et évaluer les événements du secteur de l'énergie et vue cartographiée de l'impact prévu des ouragans sur les actifs des compagnies pétrolières. L'infrastructure informatique partagée permet aux négociants d'obtenir des informations rapidement et à moindre coût. Le système fournit également un aiguillage contextuel qui aide les négociants à choisir les meilleures sources et analyses disponibles pour la tâche qu'ils souhaitent effectuer. L'expérimentation exploite la technologie MARIO (Mashup Automation with Runtime Invocation and Orchestration) pour assembler dynamiquement les applications dont les négociants ont besoin, puis déploie et exécute les composants de ces applications dédiés au traitement du flux dans un cluster InfoSphere Streams. L'ensemble de 250 analyses, sources de données et descriptions de configurations indépendantes créé pour l'expérimentation du secteur du négoce d'énergie est dynamiquement composé et paramétré en différentes combinaisons pour créer des milliers d'applications qui analysent et présentent des données associées au secteur.
- *Surveillance médicale* : le calcul des flux permet d'améliorer l'analyse médicale et de réduire la charge de travail des médecins. Des flux confidentiels de données issus de dispositifs médicaux sont analysés pour détecter des signes précoces de maladie, mettre en évidence des corrélations parmi plusieurs patients et évaluer l'efficacité des traitements. Dans ce domaine, l'accent est mis fortement sur la provenance des données et sur la détermination des mécanismes de dérivation des données lors de leur circulation dans le système. Une « première » issue de la collaboration entre IBM et une université américaine utilisera InfoSphere Streams pour surveiller les bébés prématurés dans une unité de néonatalogie.

## Présentation de l'architecture

L'architecture InfoSphere Streams représente un changement considérable de l'organisation et des capacités des systèmes de calcul. Même si elle présente certaines

similitudes avec les systèmes de traitement des événements complexes (CEP), l'architecture InfoSphere Streams supporte des débits de données plus élevés et un éventail plus large de modalités d'entrée de données. Elle fournit également un support d'infrastructure capable de répondre aux besoins d'évolutivité et d'adaptabilité dynamique (planification, répartition de la charge de travail, haute disponibilité, etc.).

Dans InfoSphere Streams, les applications continues sont composées d'opérateurs distincts qui s'interconnectent et agissent sur divers flux de données. Les flux de données peuvent provenir de l'extérieur du système ou être générés en interne dans le cadre d'une application. Le schéma ci-après (Fig. 2.13) présente la manière dont plusieurs sources de divers types de données en streaming peuvent être filtrées, classées, transformées, corrélées et/ou fusionnées pour informer des décisions dans le secteur de la finance. Ceci s'effectue à partir de calculs de revenus dynamiques ajustés en fonction d'analyses provenant d'informations sur les revenus, et à partir d'évaluations du risque en temps réel, par exemple de l'impact des dommages provoqués par un ouragan sur le point de se produire.

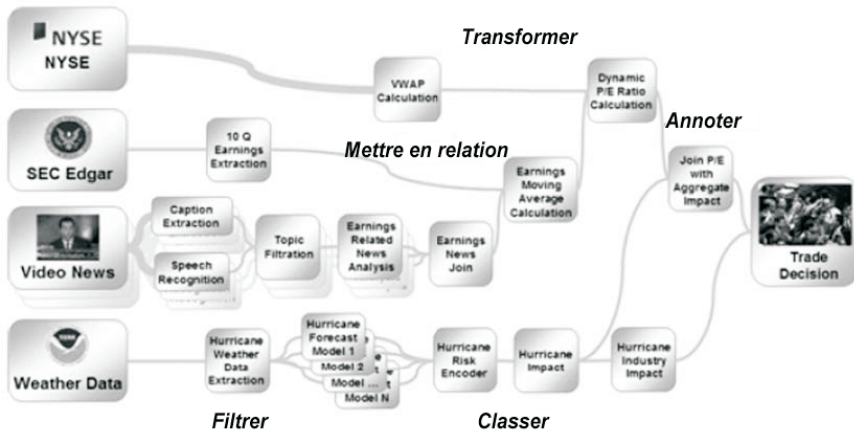


Figure 2.13 : exemple dans le secteur de la finance

Pour cette présentation, il n'est pas nécessaire de comprendre les détails de la figure 2.13. L'objectif est de montrer comment les sources de données en streaming provenant de l'extérieur d'InfoSphere Streams pénètrent dans le cœur du système, sont analysées de diverses manières par différents composants de l'application, circulent dans le système et produisent des résultats. Les résultats peuvent être utilisés de diverses façons : affichage dans un tableau de bord, actions commerciales, stockage dans des bases de données d'entreprise en vue de futures analyses, etc.

InfoSphere Streams amène le traitement de l'information à un niveau technologique radicalement plus élevé en appliquant des processus analytiques aux données en

streaming et classiques, offrant aux entreprises un avantage considérable vis à vis de la concurrence en matière de bases de données.

## Traitement analytique en ligne (OLAP)

Nous avons présenté l'analyse multidimensionnelle sous forme de feuille de calcul et sous forme d'outil graphique (Cognos). Le traitement analytique en ligne (OLAP) va encore plus loin en supportant des volumes de données beaucoup plus élevés, des analyses plus complexes et un temps de réponse extrêmement rapide. Le traitement OLAP extrait l'analyse multidimensionnelle des outils frontaux pour la placer dans l'entrepôt de données, où chaque utilisateur partage un exemplaire des données et un exemplaire de la méthode analytique. Nous avons plus ou moins tous eu l'occasion d'acheter et de financer un véhicule. Avez-vous remarqué que votre calculateur d'amortissement donne parfois un résultat de remboursement mensuel différent de celui appliqué par votre banque ? Ceci s'explique par le fait que chacun tend à utiliser les algorithmes auquel il est habitué, et que tous les algorithmes ne génèrent pas les mêmes résultats sur un même ensemble de données.

Un système OLAP peut remplacer des milliers de feuilles de calcul, chacune partageant un exemplaire des données et un exemplaire des calculs. Cependant, chaque fois qu'un système OLAP est créé, les magasins de données sont toujours déconnectés les uns des autres, et sont complexes à gérer. La figure 2.14 pourrait représenter plusieurs cubes OLAP, un cas très fréquent et facilement évitable en intégrant des fonctions OLAP dans l'entrepôt de données, comme IBM l'a fait avec InfoSphere Warehouse.

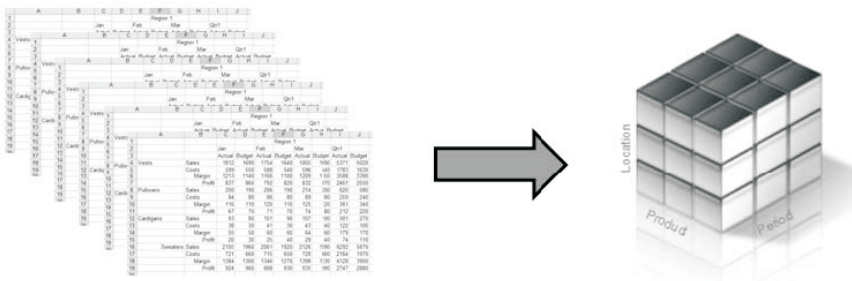


Figure 2.14 : feuilles de calcul multidimensionnelles dans un cube OLAP

L'approche d'IBM consiste à intégrer le traitement OLAP dans l'entrepôt de données pour éliminer l'ajout de magasins de données. Ceci permet d'obtenir les avantages suivants :

- Economies sur les coûts :
  - » Moins de matériel de stockage de données
  - » Coûts de licence réduits grâce à la diminution du nombre de logiciels sur moins d'ordinateurs
  - » Coûts de fonctionnement et de maintenance réduits pour certaines activités : mise à jour de logiciels, sauvegarde/restauration, collecte et synchronisation de données, transformation de données et résolution de problèmes
  - » Réduction des coûts associés aux ressources de support, notamment les coûts de formation initiale et continue (ces coûts sont particulièrement élevés dans les environnements hétérogènes)
- Mise en réseau réduite pour la connectivité et les opérations
- Meilleure productivité : la consolidation des données améliore la productivité du point de vue informatique, car la formation nécessaire concerne moins de logiciels et d'environnements matériels
- Meilleures qualité et intégrité des données : possibilité de rétablir, et même d'améliorer, la confiance de l'utilisateur vis à vis des données. La mise en œuvre de définitions de données communes et cohérentes, et de cycles de mise à jour gérés, permet d'obtenir des résultats de requêtes et de rapport cohérents parmi l'ensemble des unités opérationnelles de l'entreprise.

La notion de source de données unique pour l'ensemble des analyses fait partie des objectifs convoités par les entreprises depuis un certain temps. InfoSphere Warehouse rapproche ces entreprises du but. Une des méthodologies essentielles pour obtenir une source unique de données consiste à *intégrer l'analyse dans l'entrepôt de données*. Les services InfoSphere Warehouse Cubing Services offrent cette possibilité en exploitant les métadonnées OLAP stockées dans l'entrepôt de données, qui offre un environnement de cubes bien structuré améliorant les capacités d'analyse de l'utilisateur.

L'analyse OLAP, qui nécessitait auparavant le déplacement des données hors de l'entrepôt, produit plusieurs silos de données. L'analyse OLAP peut désormais être réalisée *sans déplacer les données en dehors de l'entrepôt*. Cette possibilité d'*analyse sans copie* représente un avantage considérable.

Le traitement OLAP, associé aux services InfoSphere Warehouse Cubing Services, permet non seulement d'obtenir une source de données unique pour l'analyse, mais également de réaliser une *analyse en temps quasi-réel* grâce à une latence réduite lors de l'accès aux données.



## Solutions IBM InfoSphere Entity Analytics

La mise en œuvre de bonnes pratiques en matière d'entreposage permet d'obtenir une seule version de la vérité, qu'il s'agisse des noms des clients ou des descriptions de produits. Mais à quel niveau de l'entrepôt se trouve la seule version de la vérité ? Les systèmes MDM (Master Data Management) permettent généralement d'obtenir cette seule version de la vérité. L'entrepôt devient ainsi une cible du processus MDM. Ceci est judicieux, au vu de la grande quantité d'énergie et de programmation nécessaire à l'élaboration d'un système MDM. Mais comment les données entrantes sont-elles traitées lorsque quelqu'un tente de dissimuler son identité, ou lorsque des informations incomplètes sont fournies ? IBM propose la solution InfoSphere Entity Analytics, qui contribue à la vérification de l'identité, même lorsque quelqu'un tente de passer au travers de la détection. IBM InfoSphere Entity Analytic Solutions (EAS) gère dynamiquement le contexte, par lequel chaque nouvelle transaction s'ajoute aux données existantes en temps réel. Le fait de placer les transactions en temps réel dans le contexte optimise la BI.

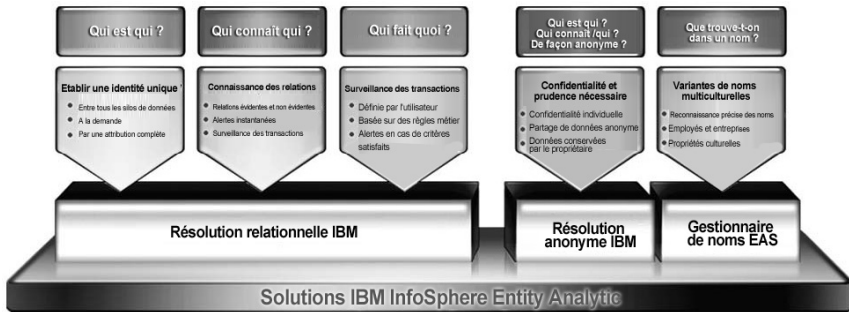


Figure 2.15 : Solutions IBM InfoSphere Entity Analytic

Observons les deux premiers composants de la figure 2.15 : Résolution de l'identité (établir une identité unique) et Résolution relationnelle (connaissance des relations).

La *résolution de l'identité* consiste à identifier les personnes et les organisations. Par exemple, dans les cercles frauduleux, il est reconnu que des individus peuvent posséder plusieurs identités : la résolution d'entité permet alors de répertorier toutes les transactions associées à chaque identité. Par exemple, EAS a permis à une entreprise de déterminer que 120 comptes clients distincts appartenaient en fait à une seule et même personne. Savoir qui est qui est fondamental à la compréhension du contexte. Un des défis majeurs auxquels font face les entreprises est l'incohérence permanente des informations sur les entités. Qui est qui ? Quel est le format des dates de naissance ? Le nom et le prénom sont-ils transposés ? Est-ce que « S. Main Street » correspond en fait à « South Maine Avenue » ? Contrairement aux techniques de gestion classiques

communes aux produits de qualité des données, EAS prend en compte qu'il n'existe pas de seule/meilleure version de la vérité lorsqu'il s'agit des données associées aux personnes et aux organisations. Les écarts potentiels offrent au contraire des indices pour les futures recherches.

La *résolution relationnelle* est le processus par lequel EAS établit qu'il existe (ou qu'il a existé) une relation entre différents individus. Par exemple :

- Deux personnes ont habité la même adresse au même moment.
- Les coordonnées de la personne à contacter en cas d'urgence d'un nouvel employé (dans le système de paie) sont celles d'un fraudeur récemment arrêté.
- Un important fournisseur a le même numéro de téléphone qu'un employé.

Cependant, chaque relation n'est pas prise en compte. Le fait que deux personnes vivent ensemble, par exemple deux clients colocataires, ne fait pas nécessairement l'objet d'une attention particulière. Bien qu'EAS ait connaissance de ces relations et les retienne, seules les relations importantes font l'objet d'alertes (par exemple, si le directeur des achats habite à la même adresse qu'un fournisseur).

EAS comprend les relations explicites (par exemple, référence sur une candidature d'embauche) et détecte les relations non explicites (par exemple, colocataires ou personnes ayant la même adresse ou le même numéro de téléphone). La compréhension du contexte sous forme de relations entre personnes apporte une visibilité essentielle pour optimiser les opportunités concernant les clients, améliorer l'acquisition et la fidélité des clients et limiter les risques.

Comment la solution EAS peut-elle améliorer davantage votre système Smart Analytics ? EAS offre une fenêtre d'observation au niveau de l'ingestion des données pour la détection en temps réel. Les nouvelles observations sont regroupées et associées aux données d'historique. Le contexte et la détection s'accumulent en temps réel. Cette visibilité est accessible aux utilisateurs. Les utilisateurs n'ont pas besoin de poser des questions compliquées pour obtenir des informations utiles issues des identités et des relations. Le système détecte des informations pertinentes en temps réel dans les flux de données, et envoie des alertes aux utilisateurs ou aux systèmes.

## Exploration de données

La technologie d'exploration de données évolue activement depuis quelques dizaines d'années. La dernière tendance des environnements de la « nouvelle intelligence » consiste à intégrer l'exploration de données dans les processus métier et les applications utilisateur. L'objectif est de permettre à l'utilisateur des domaines commerciaux non

expert en exploration de données, mais sachant comment exploiter les résultats de l'exploration, de résoudre divers problèmes courants.

L'exploration de données fait appel à des techniques statistiques et des algorithmes mathématiques évolués pour analyser des volumes de données d'historique généralement très élevés. Les objectifs de l'exploration de données consistent à rechercher et modéliser des dépendances et des comportements inconnus ou mal compris associés aux données, afin de générer des modèles descriptifs et/ou prédictifs pour créer une fenêtre d'observation et prévoir les résultats à forte valeur commerciale. Cette visibilité et la capacité à prévoir les résultats peuvent améliorer considérablement la prise de décisions avisées, pour de meilleurs services, de plus faibles coûts et une meilleure rentabilité. L'exploration de données offre un large éventail d'applications métier à forte valeur ajoutée adapté à de nombreux secteurs et scénarios. Les exemples ci-après illustrent quelques-uns des nombreux scénarios dans lesquels l'exploration de données peut apporter de manière démontrable une valeur ajoutée considérable :

- *Vente croisée, vente incitative, conception promotionnelle, accélération des recettes, fidélité des clients* : détection de segments clients (figure 2.16) et des habitudes d'achat, affinités produits et prochains achats les plus probables dans chaque segment
- *Fidélisation de la clientèle* : prévision des clients à forte valeur et à risque d'agitation
- *Gestion de la maladie* : prévision des patients présentant un risque élevé d'hospitalisation ou de déclaration d'une maladie spécifique ; progression des maladies ; interactions médicamenteuses ; choix de sujets d'étude ; résultats de traitements
- *Assurance qualité, limitation des plaintes associées à la garantie* : prévision des composants susceptibles de tomber en panne simultanément ou les uns après les autres ; recherche des causes des défaillances de composants ou des problèmes de production
- *Détection de la fraude* : détection de groupes de plaintes ou d'individus à fort potentiel de fraude
- *Gestion des catégories et des stocks* : prévision des ruptures de stock et des surstocks ; réapprovisionnements en vue de promotions sur un produit spécifique ; profilage de magasin/filiale

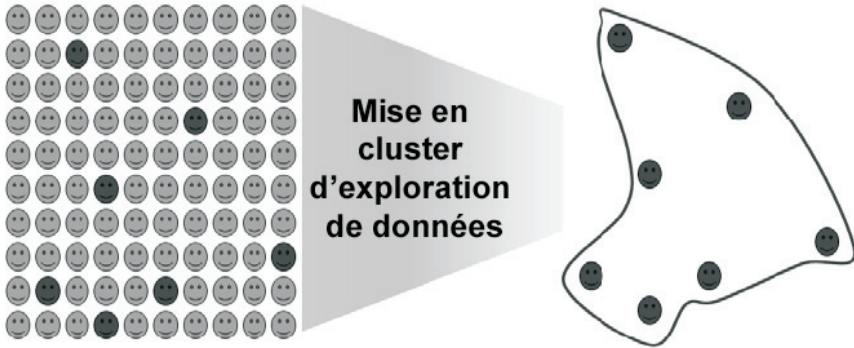


Figure 2.16 : exemple de mise en cluster d'exploration de données : détection de clients aux comportements d'achat similaires

L'exploration de données implique des techniques à la fois descriptives et prédictives. Les méthodes d'exploration descriptive (également appelée « apprentissage libre ») portent sur la mise en cluster (segmentation), les associations (analyse des liens) et les séquences (liens temporels). Exemple : segmentation de la clientèle pour détecter un ou plusieurs segments à fort potentiel sur lesquels cibler une campagne promotionnelle + associations et/ou séquences pour contribuer à la conception promotionnelle (article en promotion et articles d'affinité) + orientation des décisions de mise en stock pour anticiper l'augmentation des ventes. Les méthodes d'exploration prédictive (également appelée « apprentissage supervisé ») portent sur la classification et la régression. Exemple : prévision des manquements au remboursement de prêts à partir d'un modèle de classification en arborescence de décision, pour prévoir quels types de clients présentent un risque élevé de faillir au remboursement d'un crédit, et pour déterminer les facteurs de prévision essentiels du manquement au remboursement.

Une fois qu'un modèle d'exploration de données a pu être validé au moyen de données historiques, il peut être appliqué à de nouveaux dossiers et à des dossiers existants (dossiers client, réclamations, etc.) pour prédire des résultats, classer les clients dans le segment de marché qui leur correspond le mieux ou encore établir la probabilité d'événements ou de décisions à venir pour un individu donné, et ce via un processus de scoring, comprenez de notation statistique. Les processus de scoring peuvent être exécutés par lots (ex. rescoring programmé des titulaires d'un prêt aux fins d'actualisation des risques par défaut de chaque titulaire dans un rapport hebdomadaire) ou en temps réel (ex. scoring d'un client lors d'une transaction en ligne ou avec le centre d'appel avant validation d'un prêt ou d'une offre croisée). Le scoring peut s'effectuer au moyen d'une application utilisateur (ex. portail pour agent de centre d'appel) ou d'un processus automatisé (ex. application pour site Web de scoring automatisé pour suggérer des offres pertinentes en temps réel lorsque le client ajoute un article à son panier, mise à

jour automatisée de la propension d'un client à contracter une assurance vie en cas de changement de situation, envoi automatique d'un courrier publicitaire).

Si certains modèles d'exploration de données peuvent être adaptés à un type de problème précis d'un domaine spécifique, les approches multi-problèmes d'une solution d'exploration de données peuvent s'appliquer à tous les domaines. Aussi, la difficulté de fidélisation de la clientèle du secteur des télécommunications est théoriquement identique au risque de contracter une maladie dans le secteur de la santé. Si les paramètres sectoriels peuvent varier, le concept fondamental de prévision des résultats reste le même.

L'exploration de données peut être exploitée de différentes manières par divers utilisateurs d'une entreprise, selon les objectifs stratégiques à atteindre et l'organisation de l'entreprise. Premièrement, que ce soit à partir d'un environnement d'exploration de données et/ou des fonctionnalités d'exploration d'une base de données, les experts de l'exploration de données créent des modèles d'exploration ad hoc en réponse à certains problèmes (souvent complexes). Ces modèles peuvent très bien ne servir qu'une fois, sans être étendus à tout un groupe d'utilisateurs ou d'analyses.

Deuxièmement, un expert de l'exploration de données peut créer des modèles d'exploration de données en vue de les intégrer à des applications utilisateur, lesquelles incluront les résultats d'exploration dans des rapports ou applications interactives à l'attention des salariés côtoyant les clients ou des dirigeants.

Troisièmement, il est possible d'intégrer des fonctionnalités d'exploration de données à des applications utilisateur qui permettront aux analystes d'entreprendre des opérations d'exploration de données « guidées » et d'exploiter les résultats ainsi obtenus au sein de rapports analytiques (ex. application de mise en cluster d'un magasin de vente au détail utilisée par des responsables de produits).

Quatrièmement, les modèles d'exploration de données et le scoring peuvent être intégrés à des processus opérationnels automatisés, tels que des processus de modélisation automatisés conçus pour actualiser (mettre à jour) un modèle d'exploration de données chaque fois que de nouvelles données sont disponibles. Par exemple, les données actualisées en temps réel peuvent être exploitées par une application à même de prévoir la propension d'achat d'articles en ligne durant les Jeux Olympiques alors que des préférences se dessinent à l'occasion d'un événement populaire. Elles peuvent également être intégrées à un processus de scoring automatisé, conçu pour déclencher une action lors d'une mise à jour, comme par exemple inviter un agent d'assurance à proposer une assurance vie à son client qui vient de déclarer une nouvelle personne à charge. Enfin, elles peuvent servir à avertir des représentants commerciaux dès que certaines transactions sont effectuées, pour présenter en temps réel une offre de vente croisée à un consommateur en ligne, par exemple.

L'exploration à partir d'une base de données présente un avantage de taille sur l'exploration dans un environnement d'analyse distinct. En effet, elle offre un accès direct aux données du référentiel central et dispense ainsi de déplacer les données de la base de données vers l'environnement analytique et inversement. Dans un entrepôt de données, l'exploration de données s'effectue justement *dans* la base de données et ne nécessite aucune opération coûteuse et fastidieuse d'extraction des données vers des structures externes. Grâce à cette approche, les tâches d'exploration de données sont exécutées plus rapidement, dès le stockage des données dans l'entrepôt ou presque, en particulier lorsque les processus d'exploration sont automatisés.

L'*analyse de séries chronologiques* permet de prévoir une série de données à venir, notamment de prédire les chiffres de vente mensuels de l'année suivante. Contrairement à la plupart des techniques d'exploration de données, dont l'efficacité croît parallèlement au volume de données et qui détectent les dépendances et relations entre les champs de données, l'analyse de séries chronologiques peut générer des prévisions à partir de champs individuels et d'un volume de données nettement plus restreint (similaire à celui d'une feuille de calcul).

Pour assurer la prise de décisions réellement avisées, une solution d'exploration de données doit être intégrée à l'environnement d'analyse et ses résultats doivent pouvoir être exploités immédiatement par les utilisateurs. Pour tirer parti des résultats de l'exploration de données, il convient de développer et déployer des interfaces utilisateur, pour invoquer les fonctions d'exploration et présenter les résultats sous la forme d'aperçus ou de tables, prêts à être incorporés à des rapports, analyses, tableaux de bord et mashups composant le premier niveau de l'environnement analytique. Les outils de conception doivent quant à eux inclure des éditeurs et assistants intuitifs pour simplifier la préparation des données et les tâches de modélisation. Il faut ensuite que les interfaces utilisateur pour l'invocation des méthodes d'exploration de données soient simples d'utilisation et basées sur les standards de l'industrie comme Structured Query Language (SQL), Open Database Connectivity (ODBC), Java et les services Web. Ces interfaces doivent pouvoir s'intégrer de manière transparente aux configurations SOA (Service Oriented Architecture) et des applications en place. Elles doivent également fournir des aperçus interactifs des résultats ou les présenter dans d'autres formats appropriés pour en faciliter la lecture, l'interprétation et l'exploitation.

L'exploration de données peut également venir valoriser d'autres processus via le standard de l'industrie PMML (Predictive Model Markup Language). PMML permet l'échange de modèles d'exploration de données entre les principaux systèmes ou applications de l'entreprise, en vue d'appliquer des modèles de différentes sources aux processus métier. Il est ainsi possible, par exemple, d'importer un modèle créé via SPSS et exporté au format PMML dans une base de données DB2 et de l'intégrer à un processus de scoring appliqué à la base de données, afin de tirer pleinement parti de toutes les performances d'un grand entrepôt de données.

## Analyse textuelle

L'analyse textuelle, étroitement liée à l'exploration et à la recherche de données, émerge depuis quelques années. Traditionnellement, l'exploration de données est plus efficace lorsque les données sont numériques, structurées et ultra organisées. Or, comme nous l'avons précédemment expliqué, 80 % des données créées et la grande majorité des informations nécessaires aux processus décisionnels ne sont ni structurées, ni ultra organisées. Il s'agit de contenus textuels libres, disponibles dans une infinité de formats (XML, CSV, texte, feuilles de calcul, documents et présentations source, PDF, etc.). Ces informations n'en sont pas moins précieuses et il convient d'en analyser, explorer et révéler le contenu. C'est justement ce que permettent les technologies et méthodes d'analyse textuelle.

Les analyses textuelles visent à extraire les informations exploitables, phrases clés et indications de mesure d'une somme souvent riche de textes libres. Pour pouvoir être exploités, les analyses et rapports doivent être structurés et remis dans leur contexte : impossible de comparer ou d'ajouter deux chiffres sans la structure ni le contexte correspondants (ex. ventes de chaussures des deux derniers trimestres). Il est impossible d'envisager une structuration complète des données textuelles sachant que 80 % de nos informations sont désorganisées. Imaginez l'argent et les ressources que mobiliserait le « shredding » de chaque document XML (format semi-structuré) dans un format permettant un stockage et un accès structurés (relationnels). Imaginez maintenant faire de même avec l'ensemble des contenus textuels libres d'une entreprise. Les coûts d'un tel projet seraient astronomiques, tout comme le volume de données dupliquées. Les analyses textuelles étudient justement le texte non-structuré dans un contexte spécifique (ex. vente au détail, enquête criminelle ou culture maraîchère) et annotent ce contenu avec des métadonnées structurées. Ces annotations constituent le contenu « structuré » essentiel à l'efficacité des processus de reporting, d'analyse et de Business Intelligence.

Parmi les méthodes récentes d'analyse textuelle, la plus efficace est celle basée sur les standards UEVIA (Unstructured Information Management Architecture). Fruits de recherches universitaires et sectorielles (auxquelles IBM a largement contribué ; il est par ailleurs le premier fournisseur à avoir commercialisé des offres basées sur ces standards), les standards UIMA sont l'assurance d'un environnement d'analyse textuelle extensible. Cette capacité d'adaptation est cruciale dans un monde régi par la nouvelle intelligence.

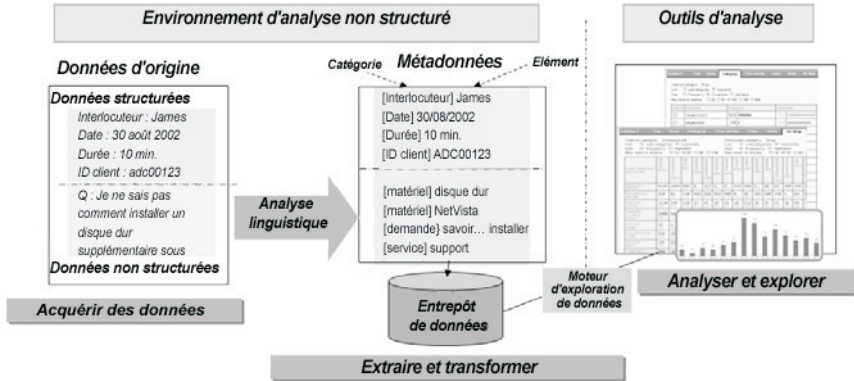


Figure 2.17 : flux d'informations d'une analyse textuelle

Comme l'illustre la figure 2.17, une analyse textuelle basée sur les standards UIMA se décompose en trois phases :

1. Ingestion de données
2. Traitement du langage naturel et indexation (annotation)
3. Visualisation et exploration interactive

L'*ingestion de données* implique d'accéder aux données dans un format commun, généralement le format XML, quelle que soit la source originelle.

La phase suivante, *de traitement du langage naturel*, est le pilier des analyses textuelles UIMA, la phase durant laquelle ont lieu les opérations de détection et d'analyse. Elle inclut toute une série de processus personnalisables :

- *Extraction du concept* : les mots-clés du concept (spécifiques à un domaine) sont repérés dans le texte source et reliés à un ensemble générique de synonymes et concepts voisins. Des dictionnaires sémantiques (personnalisables) et des règles d'appariement de formes permettent d'identifier les termes et expressions spécifiques au contexte.
- *Analyse d'intention* : il s'agit de classer les expressions grammaticales communes à l'aide des dictionnaires sémantiques et des règles d'appariement de formes. Par exemple, les verbes associés à des expressions comme « ne peut pas », « ne parvient pas à » ou « peine à » peuvent être classés dans la catégorie des « problèmes ». Ces problèmes peuvent ensuite être classés dans des groupes et



analysés. Les sujets des phrases sont ensuite liés aux problèmes, ce qui permet d'identifier un concept compréhensible : « la boucle est bouclée ».

- *Editeurs extensibles et personnalisables* : des éditeurs de dictionnaires et de règles permettent aux entreprises d'obtenir des composants libres ou vendus dans le commerce et de les étendre ou de les adapter à leurs exigences spécifiques (service client, détection de fraudes ou encore de traitement de dossiers médicaux).

La visualisation et l'exploration des résultats sont supportées par des structures d'indexation hautes performances et des composants d'exploration de texte interactifs. Le processus de visualisation fournit un aperçu graphique des annotations : fréquences et distribution du concept, distributions temporelles, tendances, anomalies des tendances et corrélations de concepts.

L'analyse textuelle UIMA est particulièrement avantageuse dans son traitement des annotations du texte. Les balises de métadonnées UIMA viennent structurer et apporter des indications de mesure aux données jusque-là non-structurées. Cela permet d'intégrer les informations de sources non-structurées à des données structurées pour les corrélérer. Il devient ainsi possible d'effectuer une analyse OLAP classique d'un cube de ventes auquel on a ajouté les commentaires en texte libre du support client. Ces nouvelles balises permettent également d'approfondir les recherches. Les concepts en texte libre auparavant indétectables peuvent à présent être indexés et interrogés, offrant ainsi une précision et une valeur ajoutée inédites.

L'analyse textuelle est une technologie dont les entreprises ne pourront bientôt plus se passer face à la diversité, au volume et à la rapidité de production des informations.

## **Gestion de contenus**

La gestion de contenus d'entreprise (Enterprise Content Management ou ECM) couvre un grand nombre de technologies et processus de gestion supportant tous types et volumes de données collectées et conservées par une entreprise. En plus de maintenir les référentiels de documents à jour, elle transforme ces ressources en informations exploitables à l'aide de processus métier et décisionnels intelligents. Tous types de contenus, pour la plupart non-structurés (environ 80 %), peuvent être gérés, y compris tous les documents professionnels, contenus Web, référentiels de messagerie électronique, communications instantanées, flux vidéo, images et graphiques, dossiers électroniques, publications, rapports, etc. Tout rapport produit par l'environnement d'analyse peut être géré par l'ECM.

Il ne s'agit pas simplement d'archiver toutes ces ressources dans un référentiel isolé. La gestion efficace des contenus d'entreprise doit rendre les informations accessibles à chaque utilisateur et processus métier impliqué dans les décisions stratégiques et la réaction aux

nouveaux événements. L'ECM valorise ces ressources en indexant les informations, en les analysant et les annotant (au moyen de l'analyse textuelle et de l'exploration de données), en les classant (par taxinomies) puis en les corrélant. Ces ressources valorisées sont ensuite mises à la disposition des utilisateurs et processus métier au travers d'interfaces basées sur les standards ouverts de l'industrie et d'API (interfaces de programmation d'applications), pour soutenir la gestion de la performance (Business Performance Management ou BPM) et les efforts d'analyse. Ces données ECM peuvent être intégrées aux applications d'analyse stratégique, de type tableaux de bord, mashups et autres applications, par lien direct et recherche approfondie. Le lien entre ces ressources, la BPM et les moteurs de règles de traitement va dans le sens de la mise en conformité aux règles en vigueur. Les technologies Web 2.0 et de collaboration permettent ensuite aux salariés de s'échanger les informations essentielles aux processus décisionnels.

La figure 2.18 résume les fondements d'une ECM efficace. Ceux-ci sont étudiés plus en détail ultérieurement, dans le cadre de l'alliance entre ECM et l'environnement d'analyse pour accélérer la prise de meilleures décisions.

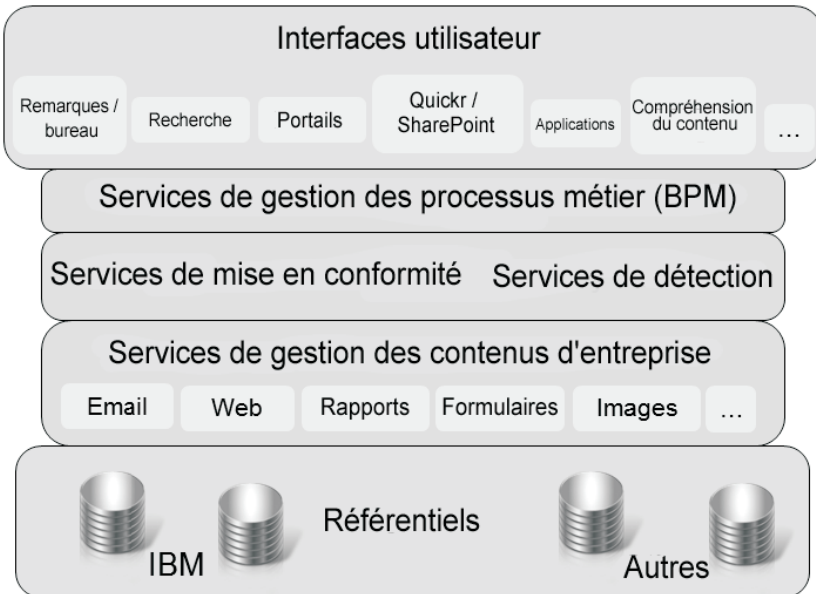


Figure 2.18 : principales composantes de la gestion de contenus d'entreprise (ECM)

- *Les bonnes informations toutes sources confondues* : les bonnes décisions se fondent sur les meilleures informations, indépendamment de la source ou du contenu. Les solutions d'ECM efficaces sont capables de gérer et de consolider les informations de divers référentiels grâce à leurs services de gestion de contenus d'entreprise.

Ceux-ci incluent des services d'accès aux contenus Web, d'e-mails, de formulaires, dossiers, rapports, images numériques, vidéos, etc. Ces services incluent également des outils de recherche et de distribution sécurisée des ressources.

- *Exploitation plus efficace des contenus* en vue d'obtenir les bonnes informations : des services de localisation et d'analyse facilitent l'analyse, l'annotation et le classement des informations. Les fonctions d'exploration de données et d'analyse textuelle décrites précédemment permettent de glaner des concepts et indications de mesure exploitables à partir des contenus gérés, indépendamment du format ou de la source, et de les annoter. Ces métadonnées intelligentes peuvent être exploitées par des processus d'analyse descendante et de BPM, ainsi que par des fonctionnalités avancées de recherche, pour optimiser les processus décisionnels. Ces services assurent également la distribution des bonnes informations en établissant des liens entre les principaux concepts extraits des informations et les processus stratégiques ou requêtes individuelles.
- *Instauration de règles et mise en conformité* : une solution d'ECM doit inclure des services de mise en conformité pour veiller à ce que les informations soient utilisées correctement et de manière sûre, conformément aux règles et politiques de confidentialité, gouvernementales et autres. Ces services facilitent la gouvernance et la maintenance des données d'entreprise tout au long de leur cycle de vie.
- *Informations fournies dans le contexte d'un processus métier, selon les règles métier* : les services de BPM permettent d'intégrer et d'incorporer les informations optimisées et conformes aux processus quotidiens pour prendre de meilleures décisions en faveur de l'entreprise. Ils permettent par exemple à un assureur traitant une déclaration de sinistre d'accéder rapidement à toutes les polices pertinentes, aux notes relatives au sinistre et à la correspondance antérieure avec le client pour guider judicieusement les décisions.
- *Distribution pertinente et collaborative des informations* : pour être réellement fructueuses, les informations doivent être délivrées à tous les utilisateurs qui en ont besoin, quand ils en ont besoin et dans le contexte requis. Pour ce faire, il convient d'utiliser des outils et interfaces dédiés, compatibles avec les outils de bureautique standard populaires comme Microsoft Excel et les navigateurs Web. Les interfaces basiques, comme les services Web et API Java, peuvent permettre d'intégrer des éléments de gestion de contenus à des applications d'analyse, portails, tableaux de bord, mashups, etc. Des liens avec des environnements collaboratifs, comme Lotus Quickr et Microsoft Sharepoint, facilitent en outre le partage efficace d'informations entre collaborateurs pour améliorer les processus décisionnels et résultats.

De nouvelles fonctionnalités d'ECM sont développées régulièrement, qui permettent aux entreprises d'exploiter toute la richesse de leurs contenus non-structurés. Outre les

simples opérations de gestion et d'archivage de ces informations, l'ECM permet d'intégrer ces ressources aux environnements d'analyse pour qu'elles optimisent les processus métier et favorisent la prise de décisions plus intelligentes.

## La couche d'infrastructure

A présent que nous avons étudié les couches de l'interface utilisateur et du traitement analytique, passons au cœur du système : la couche d'infrastructure, illustrée par la figure 2.19.



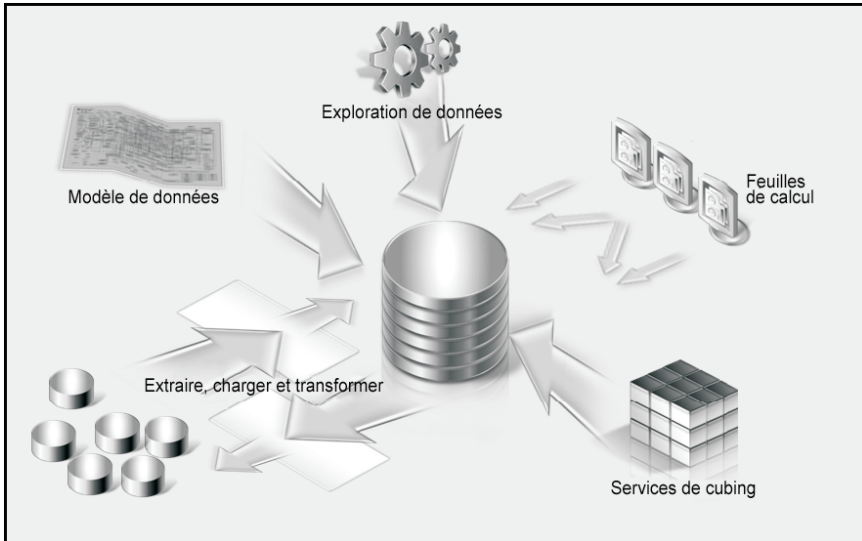
Figure 2.19 : la couche d'infrastructure

## IBM InfoSphere Warehouse

IBM InfoSphere Warehouse peut aider les entreprises à tirer parti de tous types de données ou presque en fournissant les bonnes informations au bon moment, de façon à ce que les dirigeants puissent prendre les bonnes décisions. La solution InfoSphere Warehouse allie entrepôt de données et analyse stratégique. Elle permet ainsi de définir les principaux concepts métier d'une entreprise et donne accès aux données de promotion de ces concepts au plus grand nombre dans l'entreprise. Elle aide les entreprises à extraire des données de systèmes inaccessibles par les solutions traditionnelles de BI et d'entreposage de données, permettant ainsi aux services informatiques de satisfaire les besoins d'informations exploitables, pas de simples données brutes, mais des données mises en perspective pour faciliter la prise de décisions et guider les actions.

IBM InfoSphere Warehouse est un environnement polyvalent et exhaustif qui permet aux entreprises d'accéder à des informations opérationnelles et historiques, structurées ou non, de les analyser et d'agir en conséquence. Elles disposent ainsi de la visibilité et de l'agilité dont elles ont besoin pour saisir de nouvelles opportunités, maîtriser leurs coûts et satisfaire leurs clients en permanence. Il s'agit du premier serveur de données aussi complet du marché à permettre aux entreprises d'analyser et de distribuer de manière centralisée, précise et sécurisée des informations via les applications opérationnelles et stratégiques. Contrairement aux solutions de BI et entrepôts de données traditionnels, complexes, non-intégrés et rigides, les solutions InfoSphere Warehouse simplifient la sélection, le déploiement et la maintenance d'une infrastructure de gestion d'informations

abordable, avec la flexibilité dont les entreprises ont besoin pour intégrer des données de manière dynamique et les transformer en informations stratégiques exploitables.



*Figure 2.20 : InfoSphere Warehouse*

InfoSphere Warehouse regroupe tous les composants requis pour bâtir un système d'analyse intelligent : une base de données DB2, des outils de modélisation de données physiques, ainsi que des fonctions de conversion intra-DB2 (ETL), d'ETL (extraction, conversion et chargement de données) basé sur SQL, d'exploration de données, d'analyse textuelle et d'analyse OLAP. Ces fonctions (fig. 2.20) sont intégrées par insertion de métadonnées et des outils communs viennent stimuler la productivité des développeurs et accélérer le déploiement d'applications.

## **IBM Smart Analytics System**

Comme nous ne cessons de le répéter, l'environnement économique actuel exige des entreprises capacité et rapidité d'adaptation aux conditions, défis et opportunités en constante évolution, tout en maîtrisant les coûts et en se conformant aux nombreuses réglementations de conservation des ressources. InfoSphere Warehouse constitue justement une plate-forme d'entreposage et d'analyse de données, garante de solutions de BI hautes performances. La prochaine évolution consiste à inscrire cette plate-forme logicielle dans une solution plus exhaustive encore, répondant instantanément aux exigences de nouvelle intelligence.

Alors que les changements économiques sont de plus en plus rapides, concevoir de bout en bout une solution intelligente à partir des composants de l'environnement analytique demanderait un temps et des efforts incommensurables. Les entreprises ont au contraire besoin de services et d'offres rapides à déployer, tout en profitant de l'ensemble des services à valeur ajoutée et des fonctions indispensables à un environnement intelligent. C'est exactement ce qu'offre la nouvelle solution IBM Smart Analytics System.

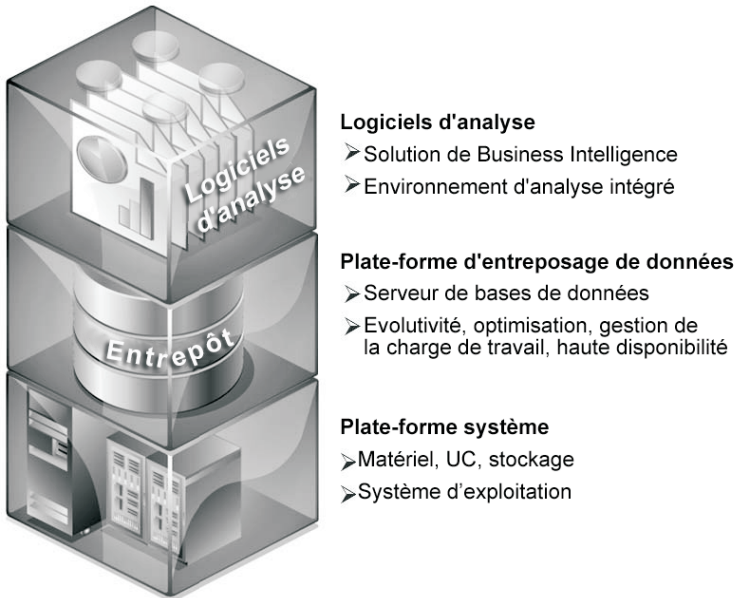


Figure 2.21 : la pile IBM Smart Analytics System

La pile IBM Smart Analytics System (fig. 2.21) est un système d'analyse hautes performances et intégré qui confère aux entreprises une visibilité en un temps record sur leurs opérations stratégiques pour aboutir plus rapidement aux meilleures décisions. Cette solution modulaire, gage d'une exceptionnelle évolutivité, s'adapte aux besoins changeants de l'entreprise et offre les avantages suivants :

- Nombreuses fonctions d'analyse
- Plate-forme robuste d'entrepôt de données
- Systèmes matériels d'IBM, parfaitement intégrés et évolutifs
- Intermédiaire unique pour le support du système ; absence de composants disparates

Ce nouveau système est préconfiguré et préoptimisé pour accélérer le déploiement de solutions de toute envergure et produire des résultats en quelques jours seulement et non plusieurs mois, contrairement aux approches d'assemblage de composants. Le principal avantage de cette solution est sa grande performance d'analyse, comme, où et quand l'exige l'utilisateur : tout ce que requiert la nouvelle intelligence. Les fonctions d'analyse intégrées d'IBM Smart Analytics System sont alignées sur l'environnement d'analyse hiérarchique décrit dans ce chapitre.

Cette solution inclut des applications avancées de reporting, de scorecards, de tableaux de bord et d'analyse pour effectuer des analyses au format et dans le contexte voulus par le processus en cours. Ces services sont basés sur la puissante plate-forme logicielle Cognos Business Intelligence et permettent aux entreprises de :

- Découvrir de nouvelles opportunités via la détection et l'exploration de données et l'analyse OLAP, par exemple
- Obtenir rapidement des réponses à leurs questions stratégiques, notamment par le biais de tableaux de bord et technologies de reporting avancés, basés sur une plate-forme conçue pour tirer pleinement parti de l'environnement d'analyse sous-jacent
- Prendre de meilleures décisions plus rapidement grâce à l'intégration hautes performances de services d'analyse, analyse textuelle et recherche incluses
- Gagner en performance grâce à la puissante plate-forme de BI et aux meilleures pratiques de Cognos

La couche de Business Intelligence (basée sur le logiciel de Cognos décrit dans le chapitre 3) repose sur l'infrastructure de l'environnement d'analyse, à laquelle elle est intégrée. Elle fournit ainsi des services d'analyse multidimensionnelle (OLAP), des outils intégrés d'exploration de données et des fonctions d'analyse de données non-structurées. Ces fonctionnalités sont intégrées directement à la plate-forme IBM Smart Analytics System et à la couche d'entreposage de données pour permettre aux utilisateurs d'effectuer des analyses hautes performances des données de l'entrepôt, sans duplication inutile ni calculs informatiques externes fastidieux. Les services d'analyse sont intégrés directement au système et les utilisateurs y accèdent de manière transparente via la couche Cognos Business Intelligence.

Toutes ces analyses exploitent un entrepôt de données d'entreprise DB2, garant d'un accès hautement évolutif et fiable aux données et informations. Les fonctions avancées de gestion de la charge de travail de l'entrepôt de données sont préconfigurées pour accélérer et optimiser les analyses.

La plate-forme logicielle est basée sur un équipement matériel d'IBM robuste et évolutif, qui bénéficie des technologies Power et AEX. Cette solution, matériel et logiciels inclus, est

entièrement modulaire. Autrement dit, les entreprises peuvent ajouter de nouveaux modules au gré de l'évolution de leurs besoins, du nombre d'utilisateurs, du volume de données, des exigences fonctionnelles et analytiques, etc. Tous les modules sont préconfigurés et optimisés. Ils peuvent donc être déployés rapidement pour répondre à de nouvelles demandes. Grâce à son architecture pleinement intégrée, IBM Smart Analytics System fournit un environnement analytique idéal sous la forme d'une solution complète, peu coûteuse à implémenter et déployer, rapide à rentabiliser et prompt à offrir des résultats.

## **IBM Smart Analytics Optimizer**

IBM Smart Analytics Optimizer est un nouveau module destiné à l'architecture IBM Smart Analytics System, une pile matérielle/logicielle révolutionnaire et exhaustive, garante de solutions de BI hautes performances. Ce module a été spécialement conçu pour accélérer les opérations d'analyse. Alliant composants matériels et logiciels, IBM Smart Analytics Optimizer peut basculer une importante partie de la charge de travail du serveur de bases de données sur un processeur distinct, améliorant ainsi à la fois la performance des requêtes traitées par le processeur et celle des requêtes restées sur le serveur.

IBM Smart Analytics Optimizer exploite de puissantes nouvelles technologies de base de données en mémoire et achemine les requêtes destinées au serveur de données vers son propre processeur. Cet acheminement et cette optimisation sont complètement transparents pour les outils et applications émettant les requêtes. Les gains de performance sont quant à eux bien visibles par les utilisateurs et processus métier qui en bénéficient. Les applications de reporting et d'analyse voient généralement leur performance multipliée par 10. L'ajout de cette technologie d'optimisation permet uniquement d'accroître le taux de réponse aux questions métier et demandes croissantes d'analyse stratégique.

## **Le point sur la nouvelle intelligence**

Ce chapitre vous a présenté les activités des analystes d'entreprise, les outils qu'ils utilisent pour orienter leurs décisions, les processus d'entrepôt de données qui supportent ces outils et l'infrastructure qui relie entre eux tous ces processus. Vous comprenez à présent que Smart Analytics n'est pas qu'un nouvel entrepôt de données et un nouveau jeu d'outils utilisateur. Il s'agit d'un système collaboratif et réactif, à même d'aider votre entreprise à pérenniser sur une planète plus intelligente.





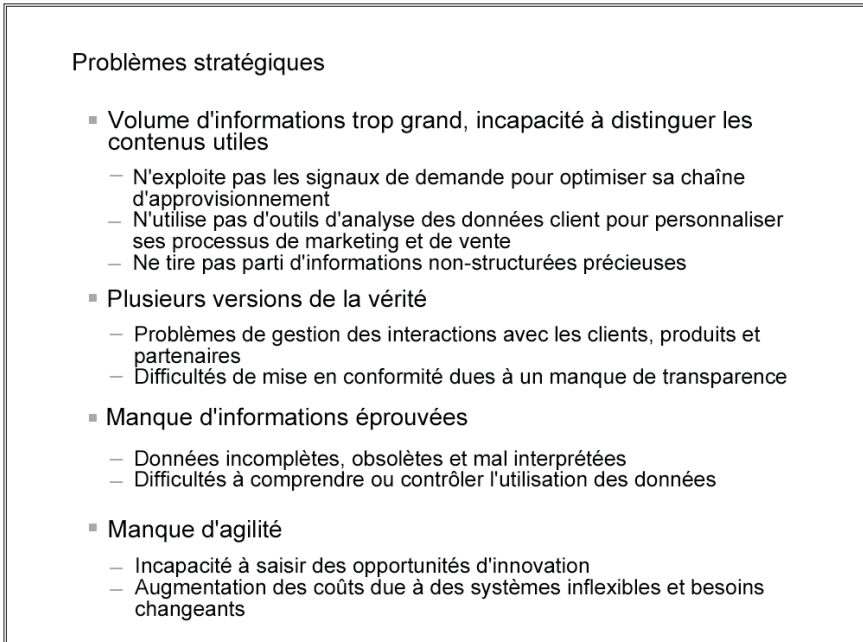
# 3

## Nouvelle intelligence Business Intelligence et gestion de la performance

**V**ous connaissez désormais les fonctions analytiques et types d'analyses que la plupart des entreprises utilisent pour améliorer la qualité des données distribuées à leurs utilisateurs. Vous savez également que les entrepôts de données et technologies d'infrastructure sont des éléments essentiels à la préparation, à la conception et à la maintenance des données et de l'infrastructure sous-jacentes. Ce que vous ignorez, c'est que les outils de Business Intelligence (BI) et de gestion de la performance valorisent l'infrastructure.

Les entreprises doivent désormais pouvoir répondre à tout un panel de questions en un temps nettement plus restreint que jamais. Les informations ne peuvent plus être cloîtrées dans des silos, ni être l'apanage de quelques-uns. Pour tirer un meilleur parti des données d'entreprise, il est essentiel de disposer d'un Information Agenda. Il faut pouvoir accéder à toutes les informations et les distribuer aux bonnes personnes dans un contexte adapté à leurs besoins professionnels. Les conditions d'accès doivent être agiles, flexibles et sans retards risquant d'affecter les processus décisionnels.

Les solutions de BI et de gestion de la performance sont utilisées depuis plusieurs années pour offrir une meilleure visibilité et permettre des analyses plus approfondies des données. Elles sont considérées depuis quelques années comme essentielles à la poursuite des objectifs stratégiques par de nombreuses entreprises. Ce chapitre décrit justement comment gagner en efficacité et mieux planifier l'utilisation et l'application de ces outils et technologies, notamment la suite Cognos 8 d'IBM.



*Figure 3.1 : problèmes communs des entreprises*

Comme l'illustre la figure 3.1, les entreprises souffrent des mêmes problèmes. L'utilisation de technologies de BI/gestion de la performance ou l'achat d'outils dédiés permet d'y remédier. De nombreux experts de l'industrie prévoient une adoption massive des solutions de BI d'ici à 2012, mais pour que cette prévision se concrétise, les entreprises vont devoir opérer plusieurs changements fondamentaux.

Nous allons ici évoquer les solutions de BI et de gestion de la performance disponibles pour chacun des problèmes énumérés à la figure 3.1. Trois d'entre elles répondent au problème de fourniture d'informations imprécises ou obsolètes. Nous entendons souvent « Nous disposons de plusieurs versions de la vérité ». Bien entendu, cela est irrationnel car il ne peut exister qu'une seule version de la vérité. Ce que cela révèle, c'est que le volume de données disponibles est trop important et qu'il est par conséquent difficile d'y détecter les données exploitables ou de distribuer les informations de manière opportune.

Généralement, l'absence d'une infrastructure de données pleinement intégrée est à l'origine du problème. Imaginons que vous ayez vendu un widget à M. John Doe domicilié au 123 Main Street, Cincinnati, OH 45202, au prix de 34,90 \$ + taxes, le 14/02/09. Les données relatives à cette transaction peuvent suivre un chemin sinueux avant de pouvoir être traitées et exploitées de quantités de manières différentes.

Au niveau granulaire, lorsqu'un agent du service client prend un appel de M. Doe ou que vous souhaitez promouvoir certains nouveaux produits auprès de vos clients, vous devez savoir quel volume d'affaires M. Doe a contracté auprès de votre entreprise. Aussi vous faut-il des informations actualisées et précises. La personne chargée des achats de widgets dans votre entreprise doit disposer d'un inventaire précis des widgets. La nécessité d'informations actualisées dépendra toutefois du cycle d'achat interne. Le service de comptabilité exigera quant à lui des données précises sur les ventes et taxes relatives aux achats de M. Doe.

Mais que se passerait-il si M. Doe n'avait pas acheté un widget, mais un autre produit mal étiqueté, sans que lui ni l'agent n'ait prêté attention aux détails de la vente ? Et si John n'était pas satisfait de son achat et décidait de se le faire rembourser ? Cette transaction et les erreurs qui y sont liées ne sont qu'une infime partie des sommes de données générées par une entreprise. Ce qui compte ici est de capturer, analyser et mettre en avant tous les aspects desdites informations aussi rapidement que possible. Isoler les utilisateurs des erreurs et appliquer les changements le plus rapidement possible ne peut se faire qu'avec une infrastructure de données dynamique, précise et extrêmement agile.

L'intégration aisée d'éléments de BI/gestion de la performance à une plate-forme SOA (Service Oriented Architecture) est essentielle pour concrétiser cette croissance. Remarquez que nous parlons ici d'une *plate-forme de BI* : or il existe justement une suite complète de fonctionnalités de BI aisément intégrables.

**Visibilité des informations au volume croissant Des solutions intelligentes et intégrées vous aident à gérer, protéger, traiter et analyser des volumes sans précédent de données structurées et non structurées —vous offrant ainsi une visibilité de vos données favorable à l'innovation et à l'optimisation de votre entreprise**

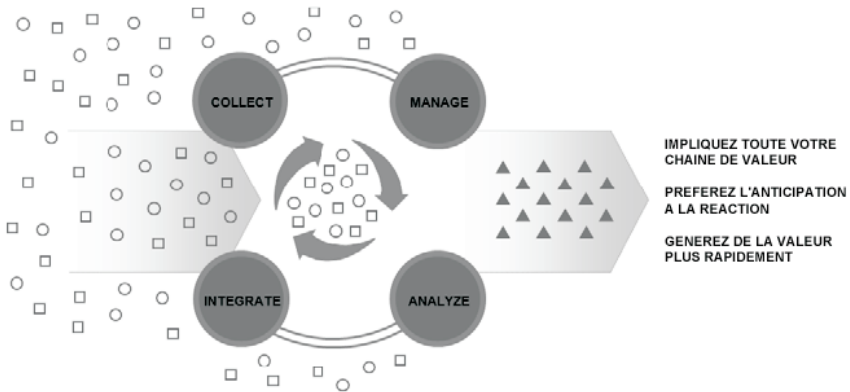


Figure 3.2 : comment exploiter intelligemment des informations dont le volume ne cesse de croître ?

Le terme « intelligent » et ses dérivés sont le leitmotiv de ce guide (fig. 3.2). En quoi les solutions de BI et de gestion de la performance vous aident-elles à opérer et agir plus intelligemment ? De nombreuses études ont été réalisées sur l'utilisation et l'efficacité des solutions de BI et de gestion de la performance, dont voici les conclusions :

1. Les solutions de Business Intelligence et de gestion de la performance constituent des investissements prioritaires pour les DSI.
2. Leur adoption par les utilisateurs reste plus lente que celle par les autres catégories d'individus de l'entreprise.
3. Une solution de BI doit être accompagnée d'un pôle de compétences, ce qui est rarement le cas, d'où une efficacité et une performance limitées.
4. La plupart des dirigeants interrogés sont conscients que leur incapacité à extraire des données pertinentes, actualisées et exploitables représente un manque à gagner.

Pour tirer le meilleur parti d'une solution de BI, il est essentiel de bien choisir les outils de BI/gestion de la performance. De l'avis du plus grand nombre, la BI devrait être facile.

La simple sélection d'un outil devrait permettre d'accéder aux bonnes données. Nous aborderons le sujet du choix des outils de BI ultérieurement. Nous nous contenterons ici de dire que le déploiement réussi de technologies de BI/gestion de la performance passe par la définition d'objectifs clairs et d'une mission stratégique précise, ainsi que par des services de support efficaces.

Quantité d'entreprises ont lourdement investi dans des outils de gestion, d'entreposage, d'analyse, d'infrastructure de données et autres afin de gagner en compétitivité et de devancer la concurrence. Certaines ont préféré les concevoir elles-mêmes.

L'implémentation de solutions de BI/gestion de la performance efficaces n'est pas garante de succès. En revanche, l'utilisation inefficace de telles technologies ou l'accès à des données non pertinentes ou obsolètes offre obligatoirement des résultats médiocres. Les solutions de BI/gestion de la performance peuvent fournir des informations essentielles à la prise de décisions intelligentes à tous les niveaux de l'entreprise et l'efficacité offerte par ces actions est non seulement quantifiable mais aussi contagieuse. Toutefois, l'application efficace des technologies de BI et de gestion de la performance repose sur de nombreux facteurs.

Commençons par la couche de données. Certaines technologies fournissent des données plus actuelles, mieux organisées et plus simples à exploiter. A l'heure de la mondialisation, de nouvelles données sont générées continuellement. Si vous pouviez les exploiter de manière plus efficace et les distribuer rapidement au sein de votre infrastructure de BI, elles pourraient vous conférer un avantage compétitif non-négligeable. Mais la BI/gestion de la performance ne concerne pas uniquement les données.

Certains facteurs déterminants affectent le succès des solutions de BI/gestion de la performance. Parmi ces principaux facteurs, découverts au fil de nos quelques années d'expérience de l'implémentation de solutions de BI/gestion de la performance, citons les suivants :

- *Les données doivent disposer d'un Information Agenda* : les données doivent être précises, actuelles et pertinentes pour pouvoir être exploitées par les processus d'analyse stratégiques. Une architecture de données de bout en bout doit être mise en place, qui pourra traiter les flux de données de manière à permettre aux entreprises de gagner en efficacité.
- *L'infrastructure doit consister en une pile complète de solutions* : l'infrastructure (ETL, entrepôt de données, accès aux données et sécurité) doit être flexible, dynamique et disponible, ainsi que sécurisée. Il faut cesser de soutenir la création de silos individuels de données et applications.
- *Les technologies de BI/gestion de la performance doivent supporter l'infrastructure sous-jacente* : ces outils doivent permettre la création et la distribution d'informations compréhensibles et exploitées par de nombreux

individus de divers services de l'entreprise. Ces outils doivent également supporter toute l'infrastructure de données. Tout processus rentrant en conflit avec les objectifs fixés pour les données et l'infrastructure de l'entreprise doivent rapidement être supprimés. Voici un exemple de processus à supprimer : la nécessité d'extraire les données après leur stockage dans l'entrepôt de données afin qu'un outil particulier puisse les utiliser plus efficacement.

- *Un plan s'impose quant à la stratégie de l'entreprise et son exécution* : il faut bien comprendre les avantages de la BI/gestion de la performance pour l'entreprise et en convaincre tous les employés amenés à utiliser ces solutions. Ce point n'est pas détaillé dans le présent chapitre et repose sur des règles et politiques internes, cependant nous maintenons qu'il est essentiel à toute entreprise.

A présent nous allons étudier la manière dont le portefeuille IBM Cognos peut profiter à tous les utilisateurs d'une entreprise. Nous évoquerons la suite Cognos en détail et décrirons comment tirer parti des solutions de BI et de gestion de la performance. Si toutes les solutions de BI incluent des fonctions plus ou moins similaires (requête, reporting, etc.), il est toutefois crucial de choisir une technologie de BI/gestion de la performance adaptée, à même de faire évoluer l'entreprise.

IBM a lancé l'initiative New Intelligence (nouvelle intelligence) pour promouvoir les solutions de BI/gestion de la performance. Notre objectif est d'aider les entreprises à prendre de meilleures décisions plus rapidement. De toute évidence, nos clients savent qu'ils ont besoin d'accéder à davantage d'informations plus rapidement. Ils semblent être conscients que l'utilisation plus efficace de technologies de BI/gestion de la performance peut leur offrir de nombreux avantages et les aider à devenir plus dynamiques. La question est de savoir comment utiliser ces technologies au mieux ? Quelques petits efforts suffiraient-ils ou est-il nécessaire d'entreprendre de véritables changements pour gagner en efficacité ? Pour commencer, demandez-vous si les systèmes et processus en place dans votre entreprise apportent des réponses à trois des questions les plus récurrentes dans votre entreprise.

## **Les trois questions essentielles au maintien de la performance des entreprises**

Comme l'illustre la figure 3.3, les solutions de BI/gestion de la performance modernes répondent aux trois questions essentielles pour comprendre une entreprise et améliorer sa performance :

## La gestion de la performance

répond à trois questions importantes pour la performance d'une entreprise

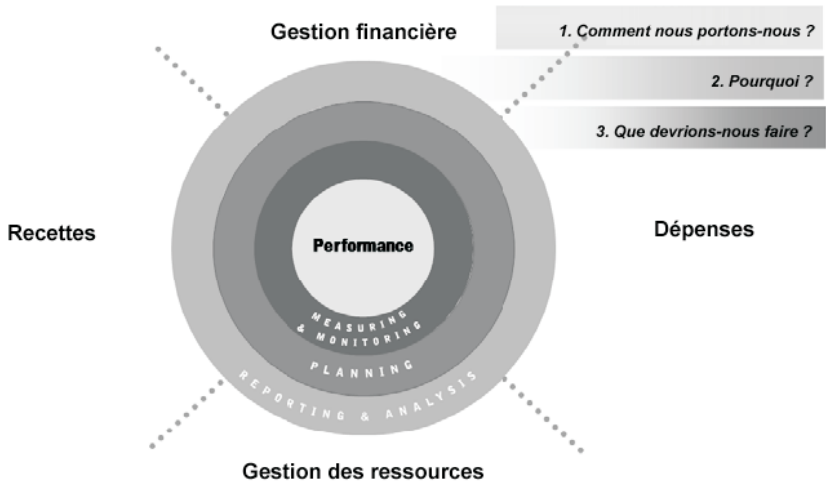


Figure 3.3 : les trois questions fondamentales

1. Comment nous portons-nous ? Exploitez des tableaux de bord et scorecards pour savoir où en est votre entreprise.
2. Pourquoi prendre ces décisions (quelles qu'elles soient) ? Servez-vous d'outils de reporting et d'analyse pour quantifier et qualifier, si possible, les indicateurs de mesure clés, tendances, positions, etc.
3. Que devrions-nous faire ? Que la situation soit satisfaisante ou préoccupante, établissez des plans, budgets, prévisions et exécutez des simulations... l'amélioration est toujours possible. Appréhendez mieux le futur.

Partout dans l'entreprise, les décideurs qui contribuent à la performance de l'entreprise doivent disposer d'outils leur permettant de trouver les réponses à ces trois questions élémentaires. Ils doivent pouvoir *mesurer et surveiller la performance* de leur entreprise et communiquer leur statut du moment. Ils sauront ainsi « Comment l'entreprise se porte ? » Pour comprendre la situation de l'entreprise, ils doivent exploiter des outils de *reporting et d'analyse qui les aideront à déterminer ce qui s'est passé*. Ils accéderont ainsi aux données historiques et pourront étudier les tendances, les anomalies et savoir comment justifier la situation.

La planification établit un lien entre ces deux réponses. Elle permet de comprendre ce qu'il se passe et ce qu'il se passera prochainement, de manière mesurable et en se basant



sur la performance effective. La planification répond à la question : *Que devrions-nous faire ?* Ces fonctionnalités intégrées vous permettent de trouver des réponses à des questions fondamentales, essentielles à l'optimisation de la performance.

Ces réponses vous aident à vous adapter aux changements dans votre entreprise et à prendre des décisions efficaces pour accroître la performance de votre entreprise. Tout décideur se pose ces questions quotidiennement. Le commerce repose sur ce cycle fondamental. L'absence d'une seule de ces réponses peut coûter très cher.

Voici un exemple de flux décisionnel : rapport trimestriel indiquant une baisse de 20 % du chiffre d'affaires par rapport à l'année précédente. Comment exploiter cette information ? Un représentant commercial demandera des rapports répondant à diverses questions commerciales : Comment expliquer cette baisse ? S'agit-il d'un problème régional ? La concurrence souffre-t-elle d'une perte similaire ? La campagne marketing a-t-elle été inefficace ? Une fois doté de ces rapports et analyses, il pourra ajuster son plan commercial et surveiller continuellement les avancées réalisées. Il s'agit d'un cycle itératif qui passe naturellement d'une question à la suivante.

## **Outils de BI / gestion de la performance : les principes de sélection pour les entreprises**

Les utilisateurs des générations précédentes d'outils et de technologies de BI se faisaient souvent la réflexion suivante : « Si seulement je pouvais accéder à mes données ! » Ce qu'ils voulaient dire, c'est que s'ils avaient eu accès aux données de tous les systèmes, silos, applications, formats et plates-formes, il leur aurait suffi d'un outil de BI pour réaliser eux-mêmes toutes les analyses requises. C'est pourquoi quantité d'outils ont été commercialisés qui permettaient d'accéder aux données de quantité de manières différentes et de les analyser. Ils incluaient une multitude de fonctions dans le seul intérêt de se démarquer de la concurrence.

Techniciens informatiques comme utilisateurs ont rapidement trouvé qu'il était bien trop complexe de disposer d'une seule version de la vérité, cohérente et opportune. Ils se sont ainsi massivement tournés vers des architectures plus hiérarchisées et orientées BI, comme les entrepôts et magasins de données. Toutefois, ces silos d'information se sont révélés moins avantageux qu'escompté tandis que l'utilisation de plusieurs outils dédiés s'est avérée coûteuse et seulement partiellement efficace.

Le temps que les entrepôts de données et solutions de BI modernes parviennent à fournir des informations utiles et actualisées, Internet faisait son apparition, multipliant le volume de données à gérer et leurs sources, ainsi que le nombre d'utilisateurs. De nouveaux standards comme XML ont émergé, engendrant une croissance exponentielle du volume d'informations disponibles. Si les technologies évoluent, les problèmes aussi.

Des défis significatifs attendent tout fournisseur souhaitant commercialiser une suite d'outils incluant toutes les fonctions requises par une entreprise actuelle. Dans le cas d'IBM Cognos, plusieurs caractéristiques distinguent ces solutions des autres offres de BI du marché :

- *Standards ouverts* : l'architecture doit être riche en fonctions et exploiter des technologies et interfaces ouvertes et acceptées dans l'industrie. Les solutions propriétaires ne séduisent plus.
- *Compatibilité avec les technologies IBM* : le portefeuille IBM Cognos supporte la riche pile technologique d'IBM et tire parti de ses nombreuses offres complémentaires (bases de données, serveurs de métadonnées, gestion de données de référence, portails, etc.).
- *Support de nombreuses solutions de tiers* : IBM Cognos n'a d'intérêt qu'associé à d'autres solutions. Ce portefeuille doit supporter d'autres bases de données, portails, solutions d'Enterprise Resource Planning (ERP), etc.

L'interopérabilité des composants d'une infrastructure est essentielle pour disposer d'une solution complète. Si chaque composant d'une solution de BI opérait de manière indépendante, rien ou presque ne différencierait une suite de solutions d'une collection d'outils de BI dédiés. En raison des nombreuses fusions et acquisitions dans le secteur de la BI ces dernières années, quantité de fournisseurs peinent à intégrer des outils disparates à une offre cohérente.

## **IBM Cognos 8 : une nouvelle suite au service d'une planète plus intelligente**

La plate-forme IBM Cognos 8 est une suite de fonctions de BI et de gestion de la performance alimentée par une couche de métadonnées qui assure la cohérence des données dans toute l'entreprise (fig. 3.4). Compatible avec de nombreuses plates-formes, depuis Windows jusqu'à System z, elle est pleinement intégrée, si bien qu'un utilisateur ou un service d'entreprise peut employer n'importe quelle fonction (requête, reporting, OnLine Analytical Processing [OLAP], performance financière, etc.) et exploiter un environnement aux fonctions communes et offrant le même aperçu des données. C'est la base pour disposer d'informations éprouvées grâce à la BI et à la gestion de la performance.

A présent que nous avons décrit les avantages et atouts des technologies et infrastructures d'entreposage de données d'IBM, découvrons plus en détail le portefeuille IBM Cognos :

1. Le portefeuille de solutions d'entreposage de données d'IBM fournit une infrastructure de base idéale pour les solutions de BI et de gestion de la performance.

2. Le concept d'une planète plus intelligente favorisant la nouvelle intelligence est basé sur l'accès à toutes les informations clés au bon moment et par les bons utilisateurs, dans le contexte de la problématique concernée.
3. La couche de métadonnées fournie aux utilisateurs et applications interagit de manière transparente avec les applications de BI et de gestion de la performance, assurant ainsi un aperçu précis, commun et synchronisé des données.

**Des solutions basées sur une SOA au service de l'efficacité de déploiement au sein d'environnements changeants**

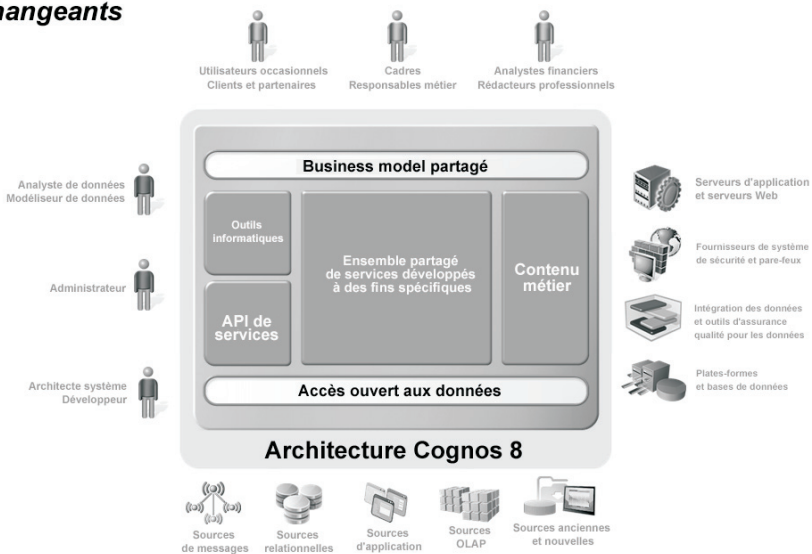


Figure 3.4 : l'architecture IBM Cognos 8, une plate-forme de BI/gestion de la performance pour entreprises intelligentes

Commençons par la couche élémentaire de l'architecture IBM Cognos 8.

### Accès ouvert aux données

Toutes les données détenues par une entreprise doivent être accessibles par la plate-forme de BI/gestion de la performance. Cognos 8 permet justement d'accéder à quantité de données de tous types, de toutes plates-formes et dans tous les formats. Les formats spéciaux, comme les cubes OLAP, doivent être disponibles, tout comme les sources de données non-structurées et semi-structurées. Les données récentes comme héritées doivent toutes faire partie de l'infrastructure de BI/gestion de la performance afin que l'entreprise puisse en tirer pleinement parti.

## **Couche de métadonnées commune**

Les définitions de données, ainsi que le remapping des données, doivent être accessibles par les bons utilisateurs, dans un contexte adapté à leur rôle et à leurs responsabilités. Sans base commune pour les informations consultées, les résultats offerts par une solution de BI/gestion de la performance ne sont pas fiables. La couche de métadonnées de Cognos est ultra personnalisable et peut ainsi être utilisée par tous les membres d'une entreprise. Les données peuvent être définies au moyen de termes compréhensibles par les utilisateurs. L'accès à celles-ci peut également être restreint à certains utilisateurs. La couche de métadonnées constitue une fenêtre sur l'Information Agenda. Cette fameuse « version unique de la vérité », souvent évoquée mais jamais concrétisée.

Une fois définies et validées, les données sont présentées aux utilisateurs sous forme de *packages*. Il s'agit de collections logiques de données éprouvées. Les utilisateurs peuvent se voir attribuer un ou plusieurs packages au besoin. Différents packages peuvent être attribués aux mêmes données afin de s'adapter aux différences de terminologie professionnelle. Les calculs les plus courants peuvent être définis dans la couche de métadonnées, fournissant ainsi une version cohérente des calculs sans exiger que chaque utilisateur crée son propre calcul.

Les utilisateurs peuvent suivre les données depuis leur source jusqu'à leur présentation. Ils n'ont ainsi plus aucun doute sur leur origine, leur validité et l'utilisation qui en a été faite. Leur intégration au glossaire Business Glossary d'IBM facilite leur interprétation dans des contextes professionnels particuliers. Sans données dignes de confiance, la BI ne peut optimiser les processus décisionnels.

## **Services dédiés**

Cette couche est au cœur de cette suite de solutions. Elle contient les fonctions de requête, de reporting, de tableau de bord, de scorecard, de détection d'événements et autres fonctions applicables par les utilisateurs aux données auxquelles ils ont accès. Par exemple, un rapport doit pouvoir être créé à partir de n'importe quelle source de données. Ainsi, il doit être possible d'accéder à un cube OLAP de la même manière que l'on accède à une source de données relationnelles.

## **Reporting et analyse**

En réponse aux questions fondamentales mentionnées précédemment, Cognos 8 inclut de nombreuses fonctions de reporting et d'analyse (fig. 3.5). Il est crucial de disposer de plusieurs méthodes pour distribuer ces informations. Car chaque utilisateur a sa propre façon d'interpréter les données. Tandis que l'un estimera qu'un rapport graphique détaillé lui est

utile, un autre privilégiera les graphiques ou tableaux. La présentation des informations à l'écran, sur imprimés, par navigateur et autres moyens doit donc être ultra personnalisable.

### **Pourquoi ?**



#### **■ Reporting d'entreprise**

- Support de plusieurs types de rapports : de production, gérés, ad-hoc, financiers, etc.
- Adaptation à toutes les sources de données
- Exécution depuis une seule couche de métadonnées
- Possibilité de personnalisation et de ciblage
- Possibilité de distribution par e-mail, portail, MS-Office, applications de recherche et périphériques mobiles

#### **■ Analyse**

- Exploration guidée de toutes les informations appartenant à l'entreprise
- Possibilité d'effectuer des analyses complexes et de modéliser des scénarios complexes de manière simple et rapide
- Extraction facilitée du sens des événements ou actions permettant d'améliorer la performance de l'entreprise
- Opportunité de passer du niveau de synthèse au niveau détaillé sans effort

*Figure 3.5 : des réponses à la question « Pourquoi ? »*

Les fonctionnalités de reporting d'IBM Cognos sont absolument inégalées. Elles offrent des résultats irréprochables dans de nombreux types et formats de comptes rendus. La couche de reporting inclut différents niveaux de fonctions adaptés aux compétences et exigences des utilisateurs. Elle fournit une interface intuitive, simple d'utilisation, accompagnée d'un jeu complet d'aides, de tutoriels et d'assistants utilisateur qui facilitent grandement les tâches de reporting.

Grâce à Cognos, il est aisé de définir des rapports et tableaux explorables, qui permettent aux utilisateurs d'obtenir des informations plus exhaustives sur un résultat particulier. La résolution des problèmes commence le plus souvent par des évaluations spécifiques. Vient ensuite l'approfondissement des recherches. L'un des principaux avantages des systèmes OLAP est qu'ils permettent d'explorer les données, de les disséquer à différents degrés.

L'option d'analyse d'IBM Cognos supporte une grande variété de sources OLAP, comme Cognos PowerPlay, Cognos TMI, Oracle/Hyperion Essbase, MS Analysis Services, SAP BW et la fonction Cubing Services (ROLAP) de la base de données DB2 d'IBM. Aussi, le respect des standards ouverts est garanti et les fonctions d'analyse étendues à tous types d'utilisateurs.

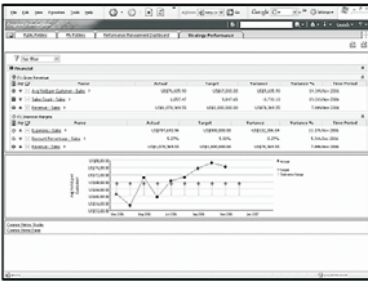
## Tableaux de bord et scorecards

### Evaluation et surveillance - Comment nous portons-nous ?



#### ■ Tableaux de bord

- Conversion d'informations complexes en présentations remarquables
- Détection des changements
- Meilleure intuitivité
- Alignement des décideurs



#### ■ Scorecards

- Evaluations instantanées liées aux cibles et bancs d'essais
- Alignement des décisions et tactiques sur les initiatives stratégiques
- Support de méthodologies de scorecarding
- Plus grande responsabilisation

Figure 3.6 : tableaux de bord et scorecards

## Détection et gestion des événements

L'efficacité d'une solution de BI/gestion de la performance réside majoritairement dans la bonne distribution des informations aux utilisateurs quand ils en ont besoin et dans un contexte adapté à leur fonction dans l'entreprise.

Certains utilisateurs ne sont concernés que par l'existence d'une valeur particulière ou l'occurrence d'un événement spécifique. Cognos 8 inclut justement une puissante fonctionnalité de déclenchement d'actions basé sur les événements, paramétrable par l'utilisateur. Il est fréquent que certains utilisateurs ne requièrent certaines informations que lorsqu'un événement spécifique se produit. Ils n'exploitent des processus de BI que rarement. Lorsqu'un événement déclenche une réponse, ils souhaitent être avertis. Cela peut être fait de différentes manières, notamment par l'envoi d'un e-mail ou d'un message sur leur PDA (assistant électronique de poche). Si l'événement ne se produit pas ou que tout va bien, ils ne souhaitent pas être dérangés.

## La couche de présentation

Comment l'utilisateur souhaite-t-il consulter les résultats des solutions de BI/gestion de la performance ? IBM Cognos 8 se base sur une architecture de clients légers mais peut fournir des résultats dans de nombreux formats et styles. Ces résultats peuvent être transférés à un navigateur, à un PDA, dans un fichier PDF ou une feuille de calcul Excel ou encore embarqués dans un processus, notamment un e-mail.

Différents utilisateurs peuvent requérir des résultats similaires dans des formats différents (fig. 3.7). L'infrastructure IBM Cognos 8 est conçue de manière à ce que les résultats, qui ne sont produits qu'une fois, puissent être communiqués dans différents formats et sous de nombreuses formes. Rien que leur présentation dans une feuille de calcul Excel s'avère extrêmement avantageuse. Elle permet d'actualiser aisément les données et résultats sous-jacents et de les transférer vers un document en conservant tous les formats et styles intacts. La feuille de calcul est ainsi autonome et n'est plus soumise aux exigences de maintien à jour des données.

### Accès flexible aux informations

- Toute langue (support robuste d'Unicode)
- Tout format (PDF, HTML, Excel, XML, etc.)
- Accès en libre-service aux alertes personnelles et à la création et exploration de données basées sur des recherches
- Distribution automatisée avec programmation, envoi d'e-mails et déclenchement d'actions basé sur les événements
- Accès depuis les applications/processus
- Accès partout où le souhaitent les utilisateurs (MS Office, périphériques mobiles, recherche...)



Figure 3.7 : accès flexible aux informations

Grâce à IBM Cognos 8, l'utilisateur peut accéder à tous types de données de l'infrastructure de l'entreprise. Il peut personnaliser l'interface en fonction de ses préférences en matière de langue, de niveau de compétences, de fonctions, de styles de résultats et de formats, et automatiser quantité de ses processus.

## **Planification, budgétisation et prévision**

Jusqu'à présent, nous nous sommes attachés à répondre à la question « Pourquoi ? ». Il va maintenant nous falloir nous pencher sur la question « Que devrions-nous faire ? ». Pour ce faire, nous vous recommandons d'utiliser Cognos Planning, Budgeting et Forecasting, qui permettent aux utilisateurs de découvrir les scénarios dont il faut tenir compte. Nous pourrions nous étendre longtemps sur le sujet du portefeuille Cognos mais nous contenterons d'évoquer dans les paragraphes qui suivent les avantages et l'impact d'une planification, d'une budgétisation et d'une prévision efficaces.

Les grandes entreprises sont régulièrement confrontées à des problèmes de planification et d'optimisation de processus. Pour ce faire, elles tirent parti des nouvelles technologies et appliquent les meilleures pratiques de l'industrie, gages d'une efficacité maximale. Elles bénéficient ainsi de plans plus précis, de prévisions et nouvelles prévisions plus actualisées et de processus décisionnels plus efficaces. Les décideurs d'entreprise se plaignent généralement des mêmes problèmes lorsqu'il s'agit de planification, de budgétisation et de prévision :

- Processus complexes et fastidieux.
- Doute sur l'intégrité des données.
- Explication des écarts difficile.
- Outils existants inflexibles et incompatibles avec un environnement dynamique.

Malgré les investissements massifs réalisés dans l'implémentation de systèmes d'ERP ces dix dernières années, la plupart des opérations de planification sont effectuées à l'aide de feuilles de calcul ou d'e-mails et mobilisent de précieuses ressources. Cette méthode s'avère coûteuse sur le long terme, les feuilles de calcul n'étant tout simplement pas conçues pour supporter ces processus de manière efficace. Voici pourquoi :

- Les règles métier (formules) sont mélangées aux données, ce qui peut engendrer des erreurs.
- Les utilisateurs s'échangent souvent des fichiers et les collaborateurs interentreprises peinent à travailler en tandem.
- Il est difficile de présenter et d'analyser les données en tenant compte de différents points de vue.
- L'agrégation de données est difficile et fastidieuse.
- Le business model n'est pas bien représenté... voire pas représenté du tout.
- Les calculs complexes, le reporting multidimensionnel et l'analyse sont pratiquement impossibles.



Comment le portefeuille IBM Cognos répond-il à ces problèmes ? Grâce à sa base de données TM1, IBM Cognos fournit un jeu d'outils et d'interfaces qui permettent aux entreprises d'isoler leurs règles, leur logique, leurs calculs complexes et leurs agrégations de leurs données. Ces solutions incluent également des modèles prêts à l'emploi, basés sur les meilleures pratiques, qui réduisent les risques et accélèrent la rentabilisation.

Les applications de planification, de budgétisation et de prévision IBM Cognos fournissent une infrastructure idéale pour les entreprises en quête d'outils de planification et de replanification rapides (fig. 3.8). Basées sur le Web, elles simplifient la collaboration à l'international.

Les applications de planification, de budgétisation et de prévision IBM Cognos permettent aux services des finances de créer et de maintenir à jour des modèles applicables à des milliers d'opérations étroitement liées. Ces modèles de planification peuvent être créés et communiqués en seulement quelques jours et non plusieurs mois. L'accélération des processus de budgétisation et de planification libère du temps que les décideurs peuvent accorder aux analyses approfondies, de manière à faire des responsables financiers de meilleurs partenaires pour l'entreprise.

### **Extension du processus de planification**

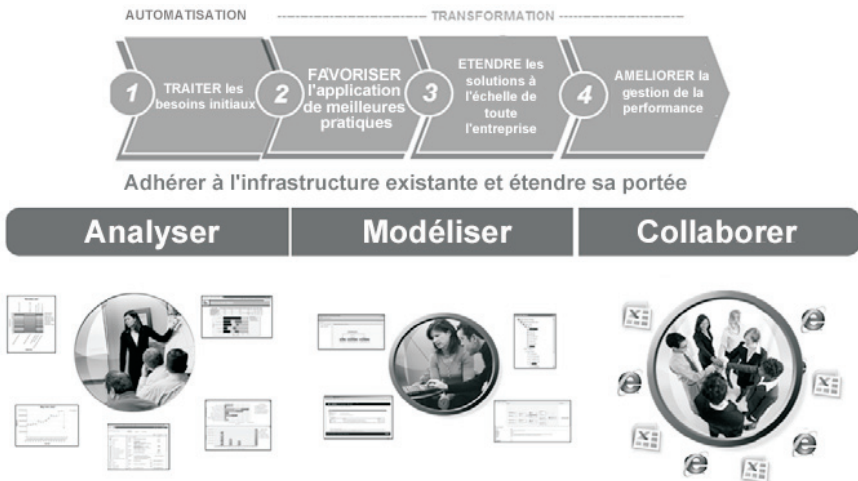


Figure 3.8 : extension du processus de planification

Le développement et la fourniture d'un processus de planification plus global permet d'accroître de manière dynamique l'efficacité de l'entreprise. Voici les avantages qu'offre la suite IBM Cognos à toutes les entreprises :

- Planification intégrée, stratégique, opérationnelle et financière depuis un même système.
- Planification collaborative en ligne qui permet la participation à tout moment et en tous lieux.
- Contrôle des versions simplifié pour des révisions fréquentes et l'exécution de simulations.
- Données actualisées en permanence, l'utilisateur pouvant ajouter ou modifier une base de données de planification centrale directement.
- Efficience accrue, les responsables des finances et directeurs de services passant moins de temps à gérer les données.
- Aperçus et formules personnalisés qui permettent d'intégrer au système les leviers de réussite de l'entreprise.
- Précision accrue grâce à un plan commun contenant moins de liens rompus, des données plus récentes et des classes mieux définies.
- Processus décisionnel géré par le service des finances.

## **Collaboration et communication efficaces**

L'une des évolutions les plus notables de la BI et de la gestion de la performance ces dernières années concerne le mode de communication des résultats et de l'impact des informations. Lorsque l'on a trouvé la solution à un ou plusieurs des problèmes qui nous affectent, comment utiliser ces informations de manière collaborative ?

IBM Cognos 8 inclut de nombreuses fonctions qui s'intègrent en toute transparence à MS Office, aux portails d'entreprise, aux mashups et autres applications. Cette suite permet aux utilisateurs de partager rapidement les résultats offerts par les solutions de BI/gestion de la performance à l'échelle internationale. Ces résultats deviennent ainsi réellement *exploitables*. Or, pour participer à bâtir une planète plus intelligente, une infrastructure doit permettre le partage rapide d'informations pour faciliter l'atteinte d'un consensus entre homologues, experts et autres collaborateurs.

Même lorsque les informations sont récentes et précises, rien ne garantit qu'elles seront interprétées correctement ou exploitées en temps voulu. L'extension d'IBM Cognos 8 à l'aide de fonctions de collaboration peut justement permettre aux salariés d'une entreprise de communiquer et de collaborer efficacement. La mise en place d'une couche de métadonnées commune assure cohérence et contrôle de toutes les informations de l'entreprise.

Le partage d'une interface commune confère également davantage de flexibilité aux membres du personnel changeant de service au sein de l'entreprise. La standardisation de l'environnement de BI de l'entreprise permet à tous les membres du personnel de l'entreprise de partager la même expérience utilisateur. La maîtrise commune de ces outils partagés permet à chacun de gagner en compétences. Le partage de fonctions communes (ex. notification d'événements et collaboration) assure également une communication cohérente entre tous les membres de l'entreprise.

## ***Interactions par portail***

IBM Cognos 8 prend en charge de nombreuses interfaces de portail. Cette suite en inclut d'ailleurs une pour les clients qui en sont dépourvus. Il est possible d'intégrer à cette interface plusieurs objets de BI/gestion de la performance à la fois. Ces objets (ex. rapports) peuvent inclure des informations de différentes sources de données et/ou de différents systèmes. Au lieu de consulter ces informations système par système, les utilisateurs disposent d'un tableau de bord réunissant toutes les informations dont ils ont besoin.

Ils peuvent également créer leurs propres tableaux de bord. Inutile ainsi d'avoir à faire appel à un expert informatique pour disposer d'un tableau de bord personnalisé. Les processus de communication gagnent ainsi en efficacité.

Divers systèmes peuvent interagir au travers de l'interface du portail. Il est par exemple possible qu'une valeur saisie dans un rapport stocké sur un système déclenche l'émission d'une requête sur un autre système. La plupart du temps, cette approche permet de collaborer et d'établir des référencements croisés sans réécrire tout le système. Elle aide ainsi à résoudre des problèmes stratégiques complexes jusqu'à présent inextricables.

## **Le dilemme : des données structurées aux données non-structurées**

Comme nous l'avons précédemment mentionné, le volume de données à gérer croît de manière considérable et rapide, au sein comme en dehors des entreprises. L'analyse efficace des données stockées sur les systèmes internes d'une entreprise, pour la plupart structurées (bases de données), relève par conséquent d'un véritable tour de force. Les utilisateurs doivent pouvoir accéder aux données des systèmes DB2, Oracle, SAP, Informix, VSAM, IMS et de quantité d'autres solutions, dans tous les formats. Sans compter les moteurs OLAP, nombreux sur le marché, qui proposent tous leurs propres magasins de données et formats, et l'émergence de nouveaux formats semi-structurés, comme XML. Toutes ces données doivent être accessibles depuis une plate-forme de BI/gestion de la performance. L'accès aux informations d'Internet est également indispensable.

Une rapide recherche sur le Web nous a permis de découvrir des statistiques étonnantes sur la génération de données. La croissance du volume de données est estimée à 15 pétaoctets par jour (1 pétaoctet correspond à 1 000 téraoctets ou encore  $10^{15}$  octets ou 1 024 téraoctets). Cette croissance phénoménale du volume de données et du nombre d'utilisateurs potentiels devrait d'ailleurs se poursuivre partout dans le monde.

Les données disponibles à l'échelle mondiale ne sont pas toutes exploitables par un processus métier. Le problème est que bon nombre d'entre elles *peuvent* s'avérer pertinentes, mais comment les détecter et en tirer parti ?

IBM Cognos 8 donne accès à des données dans quantité de formats différents, informations non-structurées incluses. Son intégration au logiciel Text Analytics d'IBM lui permet en outre de couvrir une myriade de sources de données de notes, documents, pages Web et autres.

Il est possible d'effectuer des analyses structurées sur la base d'informations non-structurées, or les solutions aux problèmes peuvent provenir d'une source restée inconnue jusqu'à cette transition, comme d'informations publiées en externe sur les derniers chiffres de vente de la concurrence, etc. L'accès à toutes les données à disposition des utilisateurs, quelle qu'en soit la source, est aujourd'hui une nécessité.

## ***IBM Cognos Go! Search : solution de recherche pour l'entreprise***

La suite IBM Cognos, fruit d'importants investissements dans le développement de technologies de recherche, est capable d'extraire des informations ultra filtrées et hiérarchisées de sources internes et externes. L'utilisation de Go! Search (fig. 3.9) permet par exemple d'interroger un référentiel de rapports connus, y compris les rapports dont on a oublié le titre ou ceux qui traitent d'un même contenu, ou simplement de consulter les rapports créés par d'autres utilisateurs.

## IBM Cognos 8 Go! Search au sein de Cognos 8 BI

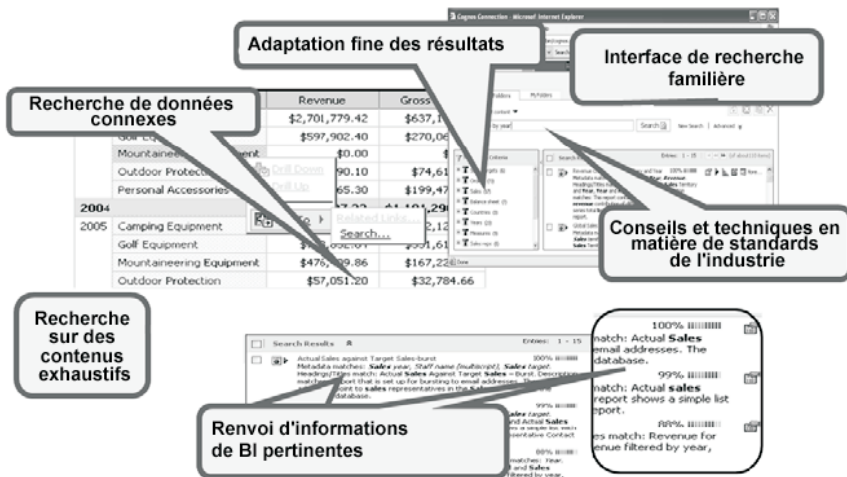


Figure 3.9 : IBM Cognos Go! Search

IBM Cognos Go! Search permet également aux utilisateurs d'interroger un gigantesque référentiel de données non-structurées (ex. Internet) pour y trouver d'éventuelles informations intéressantes ou méritant une analyse plus approfondie.

IBM Cognos Go! Search est le gage de nombreux avantages pour les clients :

- Cette offre constitue une solution de recherche pour environnements de BI préindexée, prête à l'emploi et compatible avec tous types de contenus, *qui ne nécessite aucunement l'installation d'autres produits* :
  - » 73 % des utilisateurs estiment qu'il est important de disposer d'un outil de recherche compatible avec tous types de contenus au sein d'un environnement de BI (IBM Cognos Survey).
- IBM Cognos n'a pas conclu de simples partenariats mais de véritables contrats d'intégration de technologies avec les leaders de la recherche, notamment IBM, Autonomy et Google.
- Cette solution supporte des techniques standardisées, comme les expressions types, les opérateurs booléens et l'indexation par radicaux.
- La préindexation des contenus de BI offre des résultats instantanés.
- Cette solution permet de trouver des informations contextuelles. Grâce à sa compatibilité avec tous types de contenus, les utilisateurs peuvent trouver les contenus de BI les plus pertinents.

- Elle permet également de rechercher tous types de contenus de BI, y compris les cubes, indicateurs de mesure, scorecards, événements et packages, ainsi que les rapports et tableaux de bord.

Selon la plupart des estimations de l'industrie, environ 85 % des informations électroniques sont non-structurées ou semi-structurées. Les informations internes d'une entreprise peuvent être renfermées dans des documents, images, e-mails et autres supports. A ces sources de données disponibles, il convient également d'ajouter Internet et d'autres sources de données non-structurées. Parce que toutes ces données peuvent contenir des informations exploitables par votre entreprise, Cognos vous aide à les extraire et à les interpréter.

## **La segmentation des utilisateurs au service de la Business Intelligence**

Il est terriblement naïf de croire que tous les utilisateurs peuvent exploiter des outils de BI de la même manière et ont tous des besoins et compétences similaires en la matière. Ce n'est pas pour rien que les outils de BI/gestion de la performance en libre-service connaissent un succès grandissant. La suite IBM Cognos a justement été conçue pour répondre aux besoins de tous types d'utilisateurs. Elle contient en effet une option de restriction des fonctionnalités, parfaitement adaptée aux utilisateurs les moins chevronnés. Pour les véritables experts de l'exploration de données et des analyses sophistiquées, ces fonctionnalités de base peuvent être étendues par une suite d'outils robustes.

Les outils de BI/gestion de la performance portent bien moins leurs fruits lorsqu'ils sont proposés dans le cadre d'une offre globale, qu'il n'est pas possible d'adapter à différents types d'utilisateurs. Les fonctions d'une solution de BI doivent être segmentées selon le type d'utilisateurs pour être adaptées à leurs compétences et exigences spécifiques. C'est pourquoi les fonctions et outils d'IBM Cognos 8 peuvent être sélectionnés selon les besoins de chaque utilisateur.

Dans le domaine de la BI / gestion de la performance, il est fréquent que certains utilisateurs soient chargés de communiquer les résultats aux autres utilisateurs. Ces utilisateurs disposent généralement de compétences techniques plus poussées ou de connaissances plus approfondies concernant certains processus métier, dont leurs collaborateurs peuvent tirer parti. Cognos 8 inclut de nombreuses fonctions de présentation des résultats et d'analyse approfondie qui leur permettent de présenter les résultats de BI/gestion de la performance sous la forme la plus adaptée aux destinataires.

## La segmentation des utilisateurs pour un déploiement plus efficace

Quantité d'entreprises investissent dans un grand nombre de licences individuelles pour réaliser des économies d'échelle et finissent par découvrir leurs limites en termes de capacité de déploiement. Ces limites peuvent provenir de l'absence d'un plan de déploiement ou d'une impossibilité d'adapter les fonctions aux compétences des utilisateurs de manière suffisamment granulaire.

IBM Cognos 8 a été conçue pour répondre aux besoins des différents utilisateurs en s'adaptant à leurs divers niveaux de compétences. Une entreprise compte généralement bien plus d'utilisateurs occasionnels et peu compétents que d'utilisateurs chevronnés. Toutefois, ils doivent tous pouvoir exploiter les mêmes données.

Grâce au jeu complet de fonctions d'administration de cette suite, les administrateurs peuvent sélectionner les fonctions accessibles par les utilisateurs. La présentation de l'interface utilisateur diffère ainsi d'un utilisateur à l'autre.

### Segmentation des utilisateurs et déploiement économiques

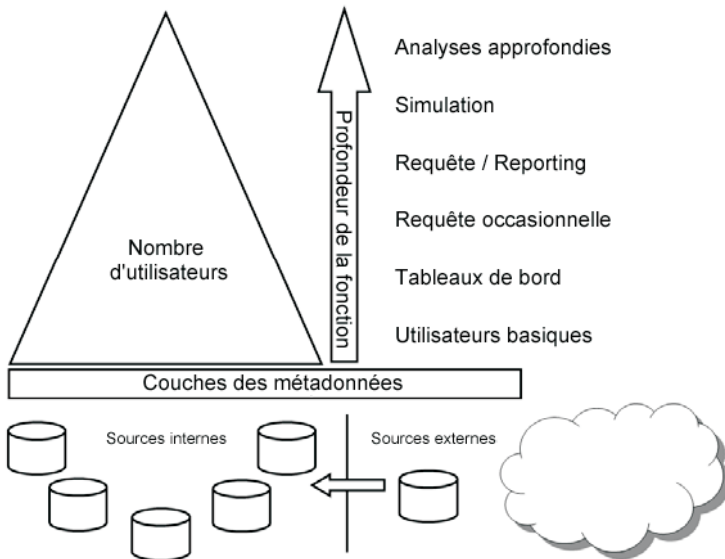


Figure 3.10 : déploiement des fonctions sur la base des besoins des utilisateurs

La figure 3.10 présente la pyramide classique des utilisateurs et leur aptitude à utiliser différentes fonctions de BI/gestion de la performance. Les catégories de fonctions ne sont pas exhaustives, mais simplement représentatives d'un type de fonctions. Ce qu'il faut retenir, c'est que cette architecture ouverte donne accès à toutes les informations requises

pour prendre des décisions avisées. Les utilisateurs se voient attribuer la juste palette de fonctions requises.

L'utilisation efficace du portefeuille de BI/gestion de la performance, ainsi que la fiabilité des données sous-jacentes, reposent sur cette plate-forme. Celle-ci améliore la confiance que les utilisateurs accordent aux données, qu'il s'agisse de simples rapports ou de résultats d'analyses approfondies communiqués par un pair.

Elle permet en outre de faire profiter le plus grand nombre des avantages de la BI/gestion de la performance en supprimant les silos de données, du moins au niveau de la couche logicielle. Les utilisateurs de l'entreprise sont ainsi segmentés en fonction de leurs compétences et peuvent accéder aux informations au moyen d'une solution de BI ouverte et centralisée.

## **Gestion des systèmes et accords de niveau de service**

IBM Cognos 8 s'accompagne d'un jeu complet de fonctionnalités d'administration et de gestion de systèmes garant d'une grande réactivité face aux changements liés aux utilisateurs et à leurs profils et d'un meilleur contrôle des ressources système. Ces fonctionnalités assurent une identification rapide et simple des problèmes de type processus qui créent des goulets d'étranglement ou mobilisent trop de ressources.

La plupart des utilisateurs d'outils et de fonctions de BI/gestion de la performance se contentent d'enquêter sur les informations qu'on leur fournit. Alors que certains se contentent de petits volumes de données, d'autres génèrent des requêtes extrêmement intensives. C'est pourquoi, au sein des environnements ad hoc, il est essentiel de bien connaître la communauté d'utilisateurs et de supporter différentes habitudes d'utilisation. Si certains systèmes de BI sont surchargés... ce n'est aucunement la faute des utilisateurs. Sans outils d'administration et de surveillance efficaces, une solution de BI est quasiment impossible à gérer. Avec les outils administratifs de Cognos, il est aisé de :

- Définir les utilisateurs, groupes, ressources
- Modifier le profil d'utilisateurs pour autoriser ou interdire l'accès à divers composants de Cognos 8
- Définir les droits d'accès aux données et/ou limites
- Surveiller l'utilisation et les ressources des systèmes
- Disposer d'un aperçu précis de l'utilisation des systèmes, de la répartition de la charge de travail, etc.



## Standardisation de la Business Intelligence

Il convient tout d'abord d'étudier de près le portefeuille de BI de l'entreprise. De nombreux clients nous disent disposer d'outils de différents fournisseurs, dont les fonctions se recourent pour la plupart. La majorité de ces offres n'ont pas été correctement déployées et, par conséquent, l'environnement de BI de l'entreprise n'est pas standardisé.

Les implications de ce scénario sont nombreuses, mais l'un des plus grands impacts de la présence d'outils disparates sur une entreprise est qu'elle réduit les opportunités de partage des compétences et résultats. Les entreprises ont acheté plusieurs outils de BI au fil du temps, développant de nouvelles expertises ici et là au sein de leur personnel. Au fil du temps, ces outils ont montré leurs limites et inconvénients, lesquels se sont multipliés au gré de l'extension de l'Information Agenda de l'entreprise.

Pour standardiser l'environnement sur une base d'outils de BI/gestion de la performance plus restreinte, un changement du mode opératoire de l'entreprise s'impose. L'une des premières étapes à suivre, et certainement la plus importante, est de mettre en place un *pôle de compétences* (CoC) dédié à la BI/gestion de la performance.

Ce CoC consiste en une organisation dont la mission est de centraliser toutes les opérations d'évaluation, de déploiement, de surveillance et de gestion d'outils, ainsi que de formation à ces outils. Il facilite ainsi l'établissement d'un environnement cohérent, garant de l'utilisation plus efficace des ressources. Les utilisateurs disposent quant à eux d'un groupe de salariés en interne dédiés à leur efficacité, qui se tiennent à leur disposition pour les aider à choisir les options les mieux adaptées à leurs besoins en matière de BI/gestion de la performance.

Lorsqu'une entreprise décide de standardiser son environnement de BI, certaines règles doivent être mises en place pour contrer les objections et faciliter l'adoption d'un nouvel outil ou d'un autre outil.

Un CoC est mieux armé pour évaluer les offres des fournisseurs et les adapter aux besoins des différents utilisateurs de l'entreprise. Les compétences et connaissances acquises en interne peuvent ensuite être partagées et appliquées à divers domaines dans l'entreprise.

Un autre facteur très important lorsque l'on envisage de standardiser un environnement de BI est de sélectionner l'outil/la plate-forme adapté(e), qui pourra offrir toutes les fonctions incluses dans les solutions remplacées. Le présent document ne fait que présenter brièvement le portefeuille IBM Cognos mais sachez que quantité d'autres produits et fonctions sont également disponibles sur les nombreux sites Web d'IBM.

Nous sommes toutefois convaincus que le portefeuille présenté ici offre à lui seul toutes les fonctions et tous les avantages dont les entreprises ont besoin.

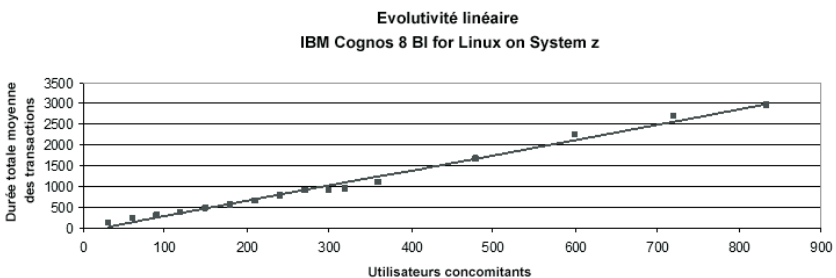
Il nous reste cependant à aborder un dernier sujet : la performance. Est-il réellement possible de déployer IBM Cognos de manière à ce qu'elle supporte des centaines, des milliers, voire des dizaines de milliers d'utilisateurs dans l'entreprise tout en offrant la rapidité, la capacité et l'évolutivité requises ?

## Performances et évolutivité

Pour s'assurer de fournir des solutions de BI/gestion de la performance PM adaptées aux entreprises, à même d'évoluer de manière linéaire et cohérente, IBM a soumis Cognos 8 à des tests et bancs d'essais rigoureux. Ces tests ont été effectués sur plusieurs plates-formes. Les résultats sont disponibles sur demande. Une simple recherche sur le Web vous permettra également de localiser les sites diffusant ces informations.

L'évolutivité linéaire est absolument indispensable aux solutions de BI/gestion de la performance efficaces et rentables. Lors de nos tests, nous avons varié les types d'utilisateurs, bases de données, volumes des requêtes et de nombreux autres paramètres système pour déterminer si Cognos pouvait respecter les meilleures pratiques de l'industrie et ainsi aider nos clients à gagner en productivité et rapidité.

***Ces tests ont confirmé qu'IBM Cognos 8 BI pouvait évoluer linéairement et prendre en charge de grands groupes d'utilisateurs.***



Le test a été mené auprès de 90 000 utilisateurs

Figure 3.11 : test d'évolutivité d'IBM Cognos 8 on Linux for System z

Le portefeuille IBM Cognos 8 prend désormais en charge Linux sur System z. Nous avons donc testé l'exécution d'IBM Cognos 8 BI sur cette plate-forme en plus des

plates-formes déjà prises en charge par son prédécesseur et avons obtenu les mêmes résultats au test d'évolutivité linéaire (fig. 3.11).

## **Etude 50TB**

Nous nous sommes intéressés à cette étude car elle est étai à l'origine destinée à permettre le développement d'une infrastructure de BI opérationnelle. La BI opérationnelle est un nouveau concept qui consiste à fournir des informations en temps réel aux utilisateurs de BI, comme les agents du service client. Si vous vous reportez à la pyramide de la figure 3.10, vous remarquerez qu'elle comporte à sa base une majorité d'utilisateurs occasionnels.

Ces utilisateurs aux compétences limitées tendent à utiliser la BI de manière intégrée et exigent des résultats précis lorsqu'ils interrogent de grands volumes de données. Lors de cette série de tests, nous avons tenu compte du fait que les utilisateurs allaient certainement interroger différentes parties de la base de données. Nous n'avons donc pas mis en cache les résultats ou requêtes de réutilisation afin de refléter fidèlement un scénario de BI opérationnelle. Comme lors de tous nos précédents tests, le « comportement » de Cognos s'est avéré cohérent et linéaire en termes d'évolutivité.

IBM exploite la suite Cognos, qui permet à ses quelque 47 000 utilisateurs de communiquer des résultats de BI/gestion de la performance à tous ses clients à l'international au moyen d'un simple jeu de services applicatifs. On nous demande souvent si nous tirons parti de nos propres technologies. Comme vous pouvez le constater, la réponse est « oui ».

## **Synthèse**

L'efficacité d'une solution de BI/gestion de la performance dépend de l'efficacité de l'infrastructure de collecte, de traitement et de transfert d'informations sur laquelle elle repose. Elle doit fournir aux utilisateurs un riche jeu de fonctions adaptées à leurs compétences et habitudes d'utilisation. C'est pourquoi nous avons développé chez IBM toute une suite de produits et de fonctions interopérables, qui s'intègrent aux plates-formes SOA modernes.

Les solutions de BI/gestion de la performance doivent établir un lien entre les données d'une entreprise et permettre le partage d'un référentiel de métadonnées commun, afin que tous les utilisateurs soient à même d'interpréter les données et leur définition, en étant certains qu'elles sont justes et à jour. Cet environnement informatique leur permet enfin de faire rapidement face aux changements et de gérer efficacement les nouvelles données disponibles. Avec Cognos 8, les techniciens informatiques peuvent aider les utilisateurs à sélectionner les outils de BI/gestion de la performance qui leur permettront de s'adapter plus efficacement à l'évolution permanente des informations et technologies.

Les salariés de l'entreprise ont également davantage confiance en l'infrastructure, les problèmes liés au volume de données à gérer, au manque de confiance et d'agilité et aux conflits entre les différentes versions de la vérité étant résolus. Cela se mesure par l'augmentation de la performance de l'entreprise et l'adoption massive du système. La confiance des utilisateurs grandit et le système s'étend à mesure que l'on constate qu'on obtient bien les informations requises dès la première tentative.

Le concept d'une « planète plus intelligente » n'est pas qu'un simple slogan marketing mais une initiative développée sur le constat que nous sommes tous, à l'échelle mondiale, plus interconnectés que jamais. Plus cette communauté internationale d'utilisateurs connectés grandit, plus les problèmes relatifs au volume d'informations et aux nouvelles sources de données se multiplient. La nécessité d'inscrire ces données dans un contexte approprié, facilitant leur interprétation, n'a par conséquent jamais été aussi grande.

Nous allons maintenant découvrir ce que doit réunir une infrastructure informatique pour supporter efficacement les applications intelligentes de demain.



# 4

## Conception de l'infrastructure informatique

**V**ous aurez beau analyser vos informations dans la volonté de découvrir tout un ensemble de nouvelles opportunités, vous hésitez à les saisir si vous manquez de confiance envers ces informations. Vos décisions deviendront alors intuitives ou subjectives, et par conséquent risquées.

### **Pouvez-vous compter sur vos informations ?**

Trouver, réunir et intégrer des informations est tout l'objectif des stratégies de la Nouvelle intelligence. Mais la valeur des informations dépend avant tout de la confiance que les utilisateurs leur portent. Se contenter de collecter des données sans comprendre leur sens ou degré de précision n'apportera que des informations difficilement exploitables par les processus décisionnels. Les entreprises disposent toutes d'informations incohérentes, incomplètes et incorrectes, exploitables dans le cadre des opérations quotidiennes mais pas à des fins d'analyse.

Quantité de projets ont échoué ou ont été grandement retardés du fait que les données étaient stockées dans un entrepôt sans que personne ne tente au préalable de pallier ces imperfections.

Les entreprises peuvent obtenir des informations fiables en mettant en place une approche systématique de l'intégration d'informations qui va bien au-delà de la simple copie de données et couvre tous les aspects de l'interprétation des informations et de la résolution des

incohérences. Chaque jour, les dirigeants des différentes divisions de l'entreprise se rencontrent pour discuter de différents sujets stratégiques, comme le chiffre d'affaires et les marges bénéficiaires. Ils disposent tous de leurs propres chiffres. Si aucun d'entre eux ne dispose de chiffres erronés, ces chiffres expriment toutefois des points de vue différents et n'offrent aucun aperçu consolidé des activités de l'entreprise. Ces différents chiffres peuvent par ailleurs se baser sur des informations non validées ou incomplètes.

Comment les entreprises s'assurent-elles que tous leurs employés sont correctement informés, impliqués et alignés sur leur processus décisionnel, de manière à ce que les décisions prises soient plus efficaces ? Sans informations correctes, les employés s'en remettent à leur intuition ou choisissent de ne pas agir.

Or, aucune de ces options n'est acceptable. Les entreprises savent qu'elles doivent mettre à la disposition de leurs utilisateurs les données fiables dont ils ont besoin quand ils en ont besoin, ce malgré leurs systèmes fragmentés et disparates, en respectant leurs exigences de volume et de rapidité, pour qu'ils aient *confiance en ces informations*.

## **Le challenge**

Les entreprises ont toujours sous-estimé la difficulté de disposer d'informations intégrées et fiables. La plupart des initiatives de gestion de la performance, d'analyse et d'entreposage de données qui tentaient de trop simplifier ou négligeaient l'intégration des informations en se contentant de copier les données ont échoué. Pour qu'une solution d'intégration d'informations garantisse un retour sur investissement satisfaisant, les utilisateurs doivent avoir suffisamment confiance en ces informations pour pouvoir agir sans hésiter et saisir les nouvelles opportunités qui se présentent. Ce chapitre traite de la manière dont les outils et fonctionnalités des offres IBM InfoSphere (fig. 4.1) peuvent aider les utilisateurs à avoir davantage confiance envers les données de leur entreprise.

## **Comment accroître la confiance des utilisateurs envers vos informations ?**

Cela passe par trois phases :

- *Compréhension des informations* via une meilleure mise en contexte et la cohérence des définitions
- *Fiabilité des informations* par la gestion de la qualité des données et son amélioration progressive
- *Accès à toutes les informations* à tout moment, grâce à de nouvelles stratégies de distribution



Figure 4.1 : le portefeuille InfoSphere—l'intégration de bout en bout

## Compréhension des informations

Le premier défi, également le plus important, est de veiller à ce que les exigences et attentes de l'entreprise soient alignées sur la structure et l'implémentation de l'infrastructure informatique. Les utilisateurs et techniciens informatiques ne parlent souvent pas le même langage et manquent des structures organisationnelles, méthodologies et technologies nécessaires pour pallier ces différences et pouvoir communiquer et collaborer efficacement.

Il est complexe de communiquer efficacement lorsque les deux parties n'emploient pas la même terminologie. Les problèmes de compréhension engendrés par l'interprétation incorrecte des informations de l'entreprise accroissent les risques d'erreurs coûteuses. Les coûts relatifs à l'ambiguïté des informations sont même plus importants.

La productivité en pâtit lorsque les cadres, en conseil de direction, ou salariés de l'atelier perdent du temps à chercher ou à tenter d'interpréter des données mal définies. Le manque de clarté des informations peut nuire aux forces de vente, faire perdre des opportunités et amoindrir la réactivité face aux évolutions du marché. A titre d'exemple, imaginons un détaillant qui déclare 125 342, 43 \$ de chiffre de vente quotidien pour son point de vente 39. Quelle est la signification de ce chiffre ? Tient-il compte des retours produit ? La réponse à cette question est cruciale et pourrait grandement affecter les décisions prises quant à ce point de vente.

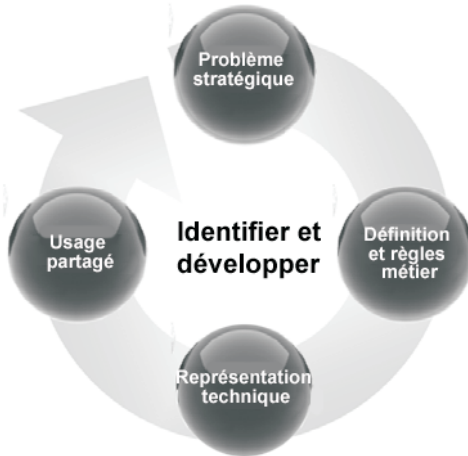


Malheureusement, il est également fréquent que les experts commerciaux et informatiques interprètent les mêmes termes métier de manière différente. Il en va de même entre les divers services d'une entreprise. Par exemple, les divers services d'une entreprise s'accorderont tous à dire qu'ils ont besoin d'informations sur les clients. En revanche, il leur sera difficile de trouver un consensus sur la définition du terme client. A quel moment une entité devient-elle un client et quand cesse-t-elle d'en être un ? Le terme « client » semble si courant que personne ne prend la peine de le définir clairement, en général, bien qu'il existe plusieurs types de clients (ex. prospects, clients actuels, anciens clients). Il s'agit de problèmes basiques, qui doivent toutefois être résolus pour que tous les acteurs de l'entreprise disposent d'un aperçu commun des données de l'entreprise.

Pire encore, le manque de précision des termes métier dissimule le désaccord fondamental sous-jacent si bien que ces débats n'ont parfois jamais lieu. Lorsque le sens de termes métier, comme les éléments d'un rapport, n'est pas clairement défini par l'entreprise, les techniciens informatiques peuvent ne pas prendre conscience de cette ambiguïté et peuvent alors prendre des décisions inappropriées quant à l'architecture informatique, basées sur des hypothèses caduques.

De nombreuses entreprises ont pris des décisions radicales pour remédier au problème d'ambiguïté de leurs données en établissant un *vocabulaire métier commun* ou un glossaire d'entreprise—un dictionnaire de référence des termes métier et relations exploités dans l'entreprise (fig. 4.2). Conçu pour convenir à tous les employés, le glossaire d'entreprise définit les termes utilisés dans les communications quotidiennes, comme les requêtes d'e-mails aux fins de rapports de business intelligence, plans marketing, documentation juridique et communications avec les clients. Ce glossaire permet également aux techniciens informatiques de veiller à ce qu'un langage approprié et commun soit utilisé par les responsables de la conception d'infrastructures, d'applications et de rapports informatiques. En d'autres termes, un glossaire d'entreprise assure la connexion des salariés de l'entreprise aux informations fiables de l'entreprise, éliminant ainsi les risques de mauvaise interprétation, source de perte de temps, d'opportunités et de revenus.

## Comprendre : définir un glossaire d'entreprise



**Impossible de se comprendre lorsque les parties en présence ne sont pas familiarisées avec la signification des termes utilisés**

Figure 4.2 : le glossaire d'entreprise

Pour résoudre les problèmes posés par l'utilisation d'une terminologie commune au sein des entreprises, celles-ci ne peuvent pas se contenter de définir des termes métier spécifiques, l'approbation des différents services métier et informatiques par rapport aux termes proposés est en effet requise. Cette terminologie commune doit également être gérée de sorte à éviter que des modifications non approuvées ne viennent involontairement en altérer le sens. A titre d'exemple, une branche d'activité spécifique de l'entreprise ne peut de son propre chef décider que les termes « prospect » et « client » sont synonymes. Par conséquent, le succès d'un glossaire unique tient essentiellement à la mise en place d'une *gestion* efficace.

Une telle gestion des données garantit en effet le recours aux structures et processus organisationnels adaptés, d'où une définition coordonnée des termes et un contrôle adéquat de leur exactitude.

Répandre l'usage du glossaire à l'échelle de toute l'entreprise *est un autre défi à relever*. S'imaginer que l'on peut résoudre le problème en regroupant un grand nombre de termes et de définitions dans un dictionnaire est illusoire : ces termes et définitions doivent être accessibles à l'échelle de toute l'entreprise et faire partie du champ lexical de tous les employés. En d'autres termes, les utilisateurs professionnels et membres du personnel informatique doivent pouvoir accéder au glossaire de l'entreprise depuis leur environnement de travail habituel. Ainsi que l'illustre la figure 4.2, les processus d'identification et de développement des termes se répètent à l'identique.

Ces processus doivent servir de base à l'élaboration d'une approche logique par rapport aux concepts métier, à leurs propriétés et à la manière dont ils sont corrélés. Dans l'idéal, le modèle logique ainsi élaboré doit présenter en détail les processus métier et autres systèmes qui comptent sur ces informations. Ce modèle doit ensuite définir en détail le contexte d'utilisation de ces données et permettre d'aboutir davantage à des définitions réutilisables, à l'échelle de toute l'entreprise, des structures informatiques.

## **Modèles de données**

Un modèle de données comportant dès le départ tous les termes métier et tous les termes techniques réduit efficacement la durée d'exécution des projets. Les modèles IBM Industry Data Models offrent un bon point de départ aux projets d'intégration des données. Des modèles de données qui définissent la manière dont les données sont structurées ainsi que les modalités à suivre pour y accéder aux fins d'analyse et de reporting sont un composant essentiel de toute infrastructure d'entrepôt de données ainsi qu'un facteur clé dans le désenclavement des silos informatiques. Les modèles de ce type contiennent des définitions pour les données techniques et métier à partir desquelles il est possible de configurer des entrepôts de données efficaces. Les entrepôts de données traditionnels sont souvent élaborés à partir de modèles de données orientés application ou développés en interne qui ne répondent pas aux exigences de l'entrepôtage en matière de données dynamiques. Cependant, l'approche développée par IBM permet aux entreprises d'élaborer un modèle de données valable à l'échelle de toute l'entreprise qui évite que les nouvelles applications ne viennent perturber les applications existantes au moment où les données sont redéfinies.

Le manque de cohérence est l'un des obstacles les plus frustrants parmi ceux qui nuisent à l'efficacité des processus décisionnels. Sans accord préalable sur les définitions et termes à utiliser pour les données, plusieurs réponses sont possibles pour une même question. Pour une compréhension des données en contexte et une prise de décision s'appuyant sur les faits, des règles et termes métier cohérents doivent être définis aussi bien pour le calcul des recettes que des coûts et des dettes.

Lorsqu'on souhaite obtenir une vision globale de l'entreprise, l'obstacle majeur qui se dresse face à une telle initiative est l'absence de modèle de données. Ce modèle est en fait un modèle unifié qui transcende tous les silos des divers systèmes opérationnels de l'entreprise. Il s'agit de d'autres termes d'un glossaire contenant les définitions et termes qui définissent l'entreprise elle-même ainsi que ses processus. Parmi les nombreuses entreprises qui ont tenté d'élaborer des modèles d'entreprise, certaines ont réussi mais beaucoup ont échoué en raison d'une insuffisante compréhension des applications et mesures clés nécessaires au fur et à mesure de leur développement. Configurer un entrepôt de données sans utiliser de modèle de données d'entreprise revient à construire une maison sans faire de plan. Mettre en

œuvrer un tel projet peut se révéler très onéreux et aboutir à un nombre infini d'itérations et de modifications afin d'intégrer les nouvelles fonctionnalités.

Ainsi que l'illustre la figure 4.3, IBM propose six modèles de données d'entreprise adaptés chacun à un secteur d'activité spécifique. Ces modèles intègrent en outre un grand nombre de modèles dimensionnels différents à même de répondre aux difficultés particulières de chacun de ces secteurs. Les modèles dimensionnels permettent une analyse approfondie d'un sujet donné, par exemple de la rentabilité.

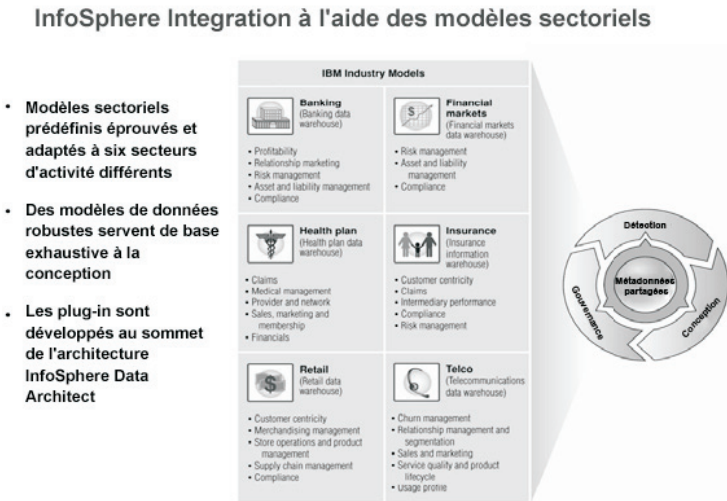


Figure 4.3 : modèles sectoriels de données d'entreprise IBM

## Profilage des données

L'analyse et la compréhension des sources d'où proviennent les informations constituent l'étape suivante. Les données se conforment-elles réellement à la définition structurelle et sémantique établie ? Existe-t-il des infractions aux contraintes d'intégrité ? Quels problèmes liés à la mauvaise qualité des données risquent de se produire lorsque les données seront intégrées ? Résoudre de tels problèmes correspond à un processus indispensable qui porte fréquemment le nom de *profilage des données*. Il s'agit en fait d'un processus ou d'une technique spécifique, généralement automatisés, qui permettent d'évaluer et de déduire statistiquement le contenu des données et les liens qui les relient entre elles à partir de données sous-jacentes réelles. Cette méthode permet également de confirmer la validité des données en fonction de règles techniques et métier prédéfinies. Les informations ainsi obtenues jouent un rôle essentiel lorsqu'il s'agit d'évaluer la viabilité des données ainsi que le niveau d'effort requis pour les nettoyer lors des étapes suivantes. La solution IBM Information Analyzer permet précisément de répondre à cette attente.

## **Traçabilité et vérifiabilité**

Une véritable compréhension des données implique une connaissance de leur origine, source ou lignée. Comment ces données ont-elles été obtenues ? Ont-elles été récupérées directement à partir d'un magasin ? Sont-elles le résultat retourné par une application comptable ? Ou proviennent-elles d'une application du service ventes utilisée pour effectuer des prévisions ? La base de données dont sont issues les données correspond-elle à leur système d'origine, c'est-à-dire au système où elles ont été saisies en première instance ou cette base de données a-t-elle été renseignée à l'aide d'une autre source de données ? L'origine première des informations, le chemin qu'elles empruntent d'une base de données à l'autre et le type d'opération qu'elles ont subies sont autant d'éléments qui influencent fortement le degré de confiance que les utilisateurs leur accordent. Il arrive fréquemment que pour une même entité apparaissant dans plusieurs systèmes, des informations divergentes soient disponibles, par exemple diverses adresses peuvent être enregistrées pour un même client en fonction du système utilisé. *L'analyse de la lignée des données* permet d'identifier leur provenance, de savoir si elles sont fiables ou non ou si elles nécessitent une vérification, un nettoyage ou une transformation supplémentaire. Identifier leur provenance ainsi que les transformations qu'elles ont subies permet aux utilisateurs d'évaluer rapidement leur degré de pertinence par rapport aux problèmes spécifiques rencontrés.

Pour une parfaite compréhension des données, les utilisateurs doivent en outre pouvoir suivre leur parcours et vérifier leur fiabilité (figure 4.4). Pour identifier toutes les modifications fonctionnelles intervenues sur les données, toutes les opérations qu'elles ont subies doivent être connues : tri, combinaison, filtrage ou calcul. A cette possibilité doit être associée, dans l'idéal, une fonctionnalité permettant aux développeurs d'analyser leur impact sur les applications qui les utilisent au cas où la définition de ces données devrait être modifiée.

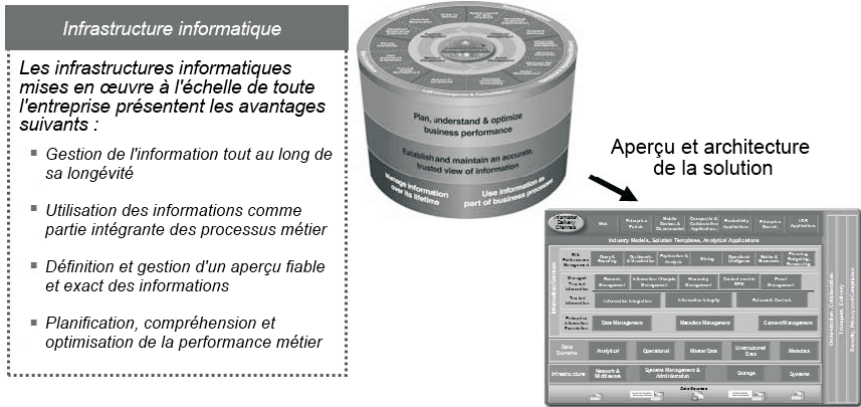


Figure 4.4 : les données doivent être traçables et vérifiables

## Inventaire de bases de données

L'identification des référentiels et bases de données contenant des informations utiles est un autre domaine nécessitant une collaboration étroite entre le service informatique et les autres services de l'entreprise. Une fois ces sources de données répertoriées, les analystes d'entreprise peuvent identifier les éléments pertinents qu'elles contiennent. Ils peuvent ensuite, au niveau supérieur, identifier la manière dont ces données sont mappées aux nouvelles structures intégrées de la solution d'analyse stratégique adoptée. Trop simples, les feuilles de calcul ne constituent pas une solution adaptée à ces mappages particuliers. Les utilisateurs ont en effet besoin d'outils plus avancés pour définir des mappages dans un format qui, sans être trop technique, leur permet de collaborer efficacement avec le personnel informatique. Bien que l'usage de feuilles de calcul soit largement répandu pour effectuer cette tâche, leur utilisation pose parfois problème.

Les feuilles de calcul n'étant pas reliés aux systèmes sous-jacents réels, il arrive fréquemment que des éléments soient manquants ou ne soient pas définis correctement. Même lorsque les informations disponibles sont justes, les copier dans des outils d'intégration nécessite souvent des opérations manuelles et mobilise à ce titre un personnel important. Les entreprises ont besoin d'un outil compatible avec le mappage des spécifications orientées métier ainsi qu'avec les artefacts techniques découlant de ces spécifications.

Enfin, comme n'importe quelle autre ressource de l'entreprise, les données doivent être inventoriées.

- Quelles données sont gérées par quel système ?
- Quels liens existent-ils entre les données appartenant à différents systèmes ?
- Comment mapper les données lorsque de nouvelles utilisations sont envisagées ?
- A quoi doivent ressembler les données pour que leur utilisation permette d'optimiser le fonctionnement de l'entreprise ?
- Comment l'entreprise utilise-t-elle les données ?
- Quelle signification ont les données pour l'entreprise ?
- Quels systèmes constituent les meilleures sources d'information pour des données spécifiques ?
- Quels sont les problèmes de qualité rencontrés au niveau des données ?

IBM Business Glossary, IBM Metadata Workbench, solution intégrée à chacun des principaux composants, et IBM Information Server offrent cette fonctionnalité. En outre, la fonction Glossary Anywhere assure l'utilisation des termes du glossaire à l'échelle de toute l'entreprise en permettant aux utilisateurs d'accéder aux définitions avalisées depuis n'importe quelle application de bureau installée sur un PC.

## ***Outils Foundation Tools***

Conçues pour aider les entreprises à configurer un environnement spécialement adapté à la gouvernance des données, les solutions Business Glossary, Information Analyzer, Industry Data Models et Metadata Workbench sont quatre exemples d'outils IBM InfoSphere Foundation Tools (fig. 4.5). Grâce à ces outils, les projets orientés information sont élaborés à l'aide d'un ensemble de solutions intégrées qui garantissent une meilleure coopération entre les équipes informatiques et celles des autres secteurs, d'où la garantie également de meilleurs résultats. Ce kit d'outils est un sous-ensemble du portefeuille de solutions logicielles IBM InfoSphere Information Server. Ce portefeuille offre les avantages suivants : assure la qualité des données au fil du temps, gère les facilités d'accès et structures afférentes, transmet aux systèmes et décideurs qui en ont besoin les données requises.

Proposant un ensemble unique de fonctionnalités, les outils InfoSphere Foundation Tools sont conçus pour faciliter la gestion et la définition des données au sein des entreprises. Ces outils aident en effet ces dernières à accomplir l'ensemble des tâches suivantes : détection et classement des nouvelles sources de données, modélisation et mappage des structures des bases de données, établissement de règles métier, mise en place d'un système d'infogérance et gestion d'un tel système, gestion de la terminologie métier et des liens hiérarchiques entre les données, enregistrement centralisé de ces informations

dans un référentiel partagé pour faciliter la coopération entre les services informatiques et les autres secteurs d'activité. Les outils IBM Foundation Tools constituent le kit élémentaire indispensable au lancement de tout projet de gouvernance des données. La gouvernance des données sera d'ailleurs l'objet du chapitre 5.

### Lancement d'un projet à l'aide des solutions Foundation Tools

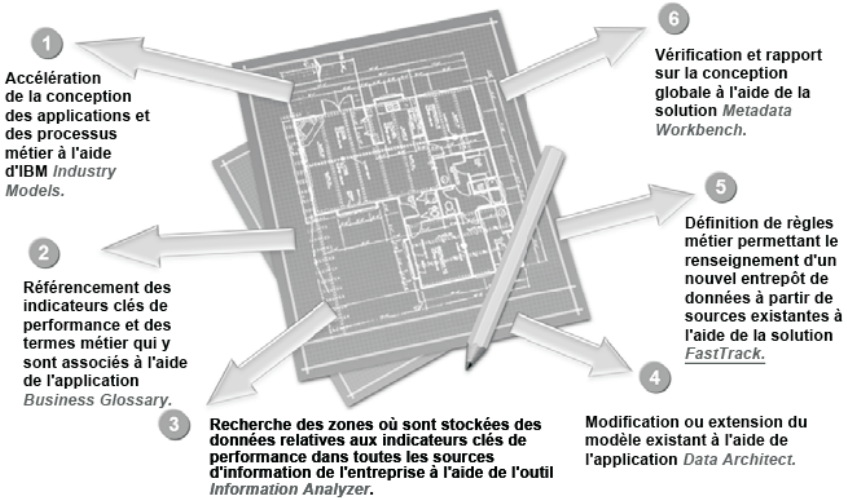


Figure 4.5 : outils Foundation Tools

Lorsqu'une stratégie de gestion des données est définie, les personnes chargées de la mise en œuvre de ce projet doivent veiller au préalable à faire coïncider leurs objectifs avec les objectifs métier des dirigeants. Le système informatique envisagé doit ensuite être compatible avec les indicateurs de mesure orientés métier de l'entreprise et pouvoir assurer le contrôle des indicateurs clés de performance. Pour garantir le respect de ces deux impératifs, les personnes chargées de la mise en œuvre doivent disposer d'un bon point de départ à même de les guider tout au long des différentes étapes de ce processus. C'est précisément ce que permettent les composants dont sont pourvus les outils IBM Foundation Tools.



## Intégration des outils InfoSphere Foundation Tools

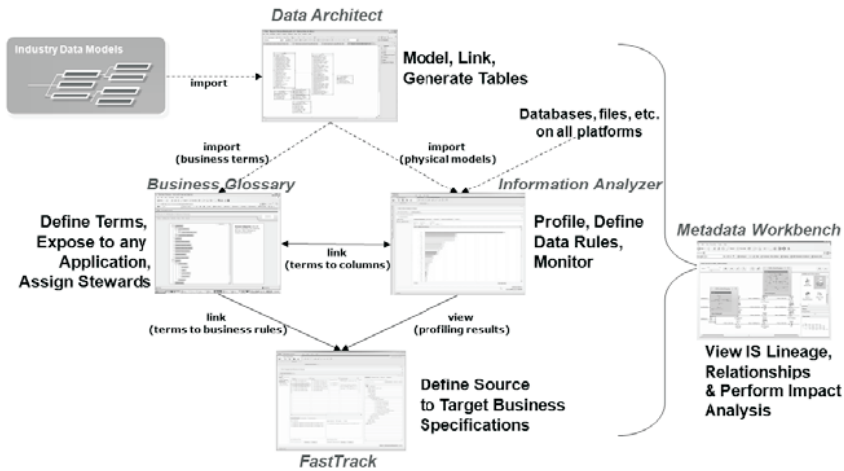


Figure 4.6 : outils InfoSphere Foundation Tools intégrés

La figure 4.6 illustre comment les fonctions de chaque outil Foundation Tools viennent compléter les fonctions de l'outil qui les précède chronologiquement tandis que leurs informations communes sont gérées à partir de l'application Metadata Workbench. Ce mode de fonctionnement autorise un fort taux de réutilisation et permet l'automatisation d'un grand nombre d'opérations manuelles qui mobilisent habituellement une large partie des ressources humaines allouées au projet.

## Fiabilité des informations

Les décisions prises ne peuvent être fiables, opportunes et efficaces qu'à la condition que les informations utilisées lors des processus décisionnels bénéficient de la confiance des utilisateurs quant à leur exactitude. Les données sont considérées comme fiables lorsqu'elles sont correctes, exhaustives et cohérentes à l'échelle de toute l'entreprise. Les informations, pour permettre un recoupement efficace, doivent respecter un certain nombre de critères élémentaires, entre autres les critères suivants : alignement des clés, disponibilité des informations stratégiques et cohérence des données figurant dans les champs. Les problèmes de qualité en matière de données, en particulier lorsqu'ils ne sont pas détectés, réduisent considérablement les chances d'obtenir, à partir des données disponibles, des réponses non erronées. Dans les cas les moins graves, de tels problèmes affectent la valeur des informations fournies, celles-ci étant inexactes ou incomplètes. Dans le pire des scénarios, les informations obtenues sont de qualité tellement médiocre qu'elles sont totalement inexploitables.

## **Profilage des données**

Dans le cadre d'un projet orienté information, tel qu'indiqué précédemment, profiler les données sources à l'aide de la solution Information Analyzer est une méthode efficace qui assure leur compréhension. Pour garantir leur fiabilité, ces données doivent être analysées et corrigées. Les champs doivent contenir les informations adéquates et toutes les clés de verrouillage avoir été validées. A ce stade du processus, automatiser autant que possible les tâches à effectuer, réutiliser les informations précédemment capturées et traiter tous les éventuels problèmes détectés constituent le principal objectif.

## **Nettoyage des données**

Pour résoudre les problèmes de qualité identifiés lors de la phase de profilage, les données doivent être nettoyées ou corrigées. Un grand nombre d'entreprises reconnaissent que les informations dont elles disposent présentent des problèmes de qualité. A moins de résoudre ces problèmes, les applications de gestion de la performance de l'entreprise ne seront pas en mesure de fournir des informations exactes. Prenons à titre d'exemple le cas d'une entreprise qui souhaite analyser la rentabilité client de plusieurs gammes de produits dont les données respectives n'ont fait jusque là l'objet d'aucune harmonisation. Si les données des clients (par exemple, leur adresse) ne sont pas nettoyées et standardisées, des informations erronées ou présentes en double risquent d'altérer négativement le résultat. Si un même client est référencé sous deux adresses distinctes dans la base de données de deux gammes de produits différentes, ce client ne pourra être identifié comme étant un seul et même client, d'où l'obtention, après analyse, de résultats erronés, notamment le fait que ces deux gammes de produits n'ont apparemment aucun client en commun. La nécessité de *nettoyer les données* implique également leur standardisation (par exemple, définir un même format pour toutes les adresses), l'ajout de nouvelles informations (par exemple, ajouter aux adresses les codes postaux manquants) et la suppression des données figurant en double. Assurer en continu la qualité des données est un processus qui implique les tâches suivantes : profilage, standardisation, appariement et surveillance. Or, ce sont autant de tâches que la solution IBM Information Server permet précisément d'accomplir.

Pratiquement toutes les études menées afin d'identifier l'origine des défaillances des entrepôts révèlent que l'une des trois principales causes d'échec lors de la mise en œuvre d'un projet d'entrepôt est généralement la qualité insuffisante des données. Mais comment est-il possible que les données utilisées par des systèmes opérationnels soient aussi imparfaites ? Trois facteurs expliquent ce phénomène :

1. Les données sont optimisées dans le but d'accomplir une tâche spécifique.
2. La qualité des données est suffisante par rapport à l'objectif particulier qu'elles permettent d'atteindre.

3. Le contexte de création des données diverge de leur contexte d'utilisation.

Les données collectées dans un système opérationnel le sont à des fins spécifiques. De nombreuses bases de données contiennent des champs de renseignements supplémentaires qui peuvent être utiles à l'entreprise, mais qui ne sont pas utilisés dans les processus quotidiens. Il en résulte que les valeurs contenues dans ces champs ne sont ni gérées ni surveillées aussi étroitement que les autres données nécessaires au fonctionnement de l'application directement concernée et elles sont donc souvent impossibles à justifier, inexactes, voire arbitraires. Il arrive également que de tels champs soient réutilisés pour résoudre des problèmes spécifiques, d'où la présence d'informations ne correspondant absolument pas à l'intitulé de ces champs.

La deuxième raison mentionnée expliquant la médiocre qualité des données correspond au fait que la qualité des données est généralement suffisante pour répondre aux besoins spécifiques d'une application donnée, mais insuffisante pour permettre une utilisation de ces données à une plus grande échelle. Les noms et adresses des clients peuvent être des informations suffisantes pour leur faire parvenir un avis par la poste, mais insuffisantes pour fusionner facilement les différentes relations qui unissent ces clients à l'entreprise.

Pour que l'entrepôt contienne des données fiables, les entreprises doivent analyser, comprendre et profiler les données sources qui y seront chargées. Les termes doivent également être standardisés et l'ensemble des données fusionnées dans un format compatible avec tous les outils et applications qui devront les utiliser. Le processus d'entreposage doit également conserver une trace de l'emplacement original des données ainsi que des transformations qu'elles ont subies. Néanmoins, la fiabilité des données ne pose pas uniquement des problèmes de nature statique. Par exemple : le service postal américain explique que plus de 40 millions d'Américains déménagent chaque année, ce qui le contraint à une mise à jour annuelle d'environ un sixième des adresses enregistrées dans sa base de données client.

Les données saisies doivent être nettoyées pour identifier celles qui sont inexactes, redondantes ou concordantes, puis corrigées, standardisées et réconciliées. Standardisation, validation, appariement et fusion sont autant d'opérations effectuées par les fonctionnalités de nettoyage sur les données qui garantissent leur vérifiabilité, cohérence, qualité, exhaustivité et fiabilité. En améliorant la qualité des informations, la solution IBM Information Server fait d'une pierre deux coups. Elle favorise en effet la confiance des utilisateurs, simplifie les processus décisionnels et améliore leur efficacité, permet un traitement plus personnel des clients, facilite l'identification des opportunités commerciales financièrement intéressantes et rassemble les informations fiables et vérifiables nécessaires au respect des réglementations telles que Sarbanes-Oxley Act et Basel II.

## ***Intégration et transformation des informations***

Dans la plupart des cas, lorsque des initiatives d'entreposage et de gestion de la performance de l'entreprise sont prises, les informations disponibles ne sont pas utilisables en l'état. Elles se trouvent en effet réparties entre plusieurs systèmes hétérogènes, dans différents formats et contiennent souvent des valeurs contradictoires. Même lorsque leur qualité n'est pas un problème, les besoins des environnements analytiques étant très différents de ceux des environnements opérationnels, les informations métier doivent être extraites dans un format adapté aux applications analytiques. Parvenir à gommer ces différences est l'un des défis les plus importants posés par l'intégration des informations. Comment les informations existantes peuvent être utilisées, transformées, puis intégrées de sorte que les utilisateurs de l'entreprise puissent en disposer dans un format adapté à leurs besoins spécifiques ? Il est à noter que les anciens systèmes ont souvent été conçus, puis commercialisés à une époque où leurs besoins étaient complètement différents. Or, ces systèmes jouent toujours à l'heure actuelle un rôle important, puisqu'ils continuent de gérer certaines des données les plus cruciales de l'entreprise. Dans la plupart des cas, les remplacer purement et simplement, et ce uniquement parce que les informations nécessitent d'être considérées sous un autre angle ou associées différemment, n'est pas la panacée. Disposer de fonctionnalités de *transformation* et de *nettoyage des données* est indispensable pour les extraire des sources d'information, les transformer, les agréger, les nettoyer, puis les charger dans un entrepôt utilisable par les applications métier.

Les importants volumes de données à traiter (toutes les données financières d'un grand groupe international, par exemple) ainsi que les délais relativement courts impartis peuvent en outre venir compliquer ce processus de transformation. Extraire les informations des systèmes opérationnels existants a souvent un impact négatif sur la performance. Les données doivent donc être transférées rapidement au cours de périodes de maintenance de plus en plus courtes. Après extraction des sources d'information, tandis que les données sont nettoyées, des règles de nettoyage spécifiques doivent être appliquées pour résoudre les éventuels problèmes de qualité.

Traditionnellement, les processus d'intégration des informations se limitent à des traitements par lots des données, lesquels se déroulent pendant les opérations de maintenance, par exemple pendant le week-end. La fréquence à laquelle interviennent ces opérations d'intégration limite le degré d'actualité des informations dont les utilisateurs de l'entreprise peuvent disposer via les applications de business intelligence (BI). A l'heure actuelle, les décisions et mesures requises doivent être prises de plus en plus rapidement au sein des entreprises. En outre, celles-ci ont même de plus en plus souvent besoin de processus de transformation et d'intégration des données pouvant s'exécuter *en temps réel*. Il est à noter toutefois que tous les scénarios d'utilisation des données ne nécessitent pas forcément une intégration en temps réel de celles-ci. Par conséquent, la technologie d'intégration utilisée doit supporter différents mécanismes de temporisation adaptés aux différentes tâches à effectuer.

Dans certains cas, intégrer les informations en se contentant de les copier dans un référentiel partagé n'est ni possible, par exemple à cause des réglementations externes ou internes en vigueur, ni faisable pour des raisons pratiques. La seule option d'intégration possible pour les informations est alors un processus *virtuel à la demande*. Dans un tel cas de figure, les informations sont extraites des sources concernées, intégrées, puis renvoyées uniquement sur demande, et ce sans être à aucun moment stockées.

Solution d'intégration des données en temps réel ultra performante à faible latence, IBM InfoSphere Change Data Capture permet aux utilisateurs d'identifier rapidement les modifications qui interviennent sur les données métier qui les concernent à l'échelle de toute l'entreprise et de répondre à ces changements de manière adaptée. Grâce aux solutions d'intégration des données en temps réel désormais disponibles, les entreprises peuvent aujourd'hui prendre des décisions commerciales plus avisées, diriger des opérations plus efficaces, accroître leur base clientèle, conclure de nouveaux partenariats et augmenter leur résultat net. La solution IBM InfoSphere Change Data Capture permet notamment à ces entreprises d'accomplir les tâches suivantes :

- Chargement en temps réel des données dans les entrepôts, d'où la possibilité pour les entreprises de prendre des décisions stratégiques et opérationnelles s'appuyant sur des informations à jour
- Acheminement dynamique des données, en fonction du contenu, jusqu'à diverses files d'attente de message utilisées par une ou plusieurs applications, garantissant ainsi l'accès à des données fiables et précises à l'échelle de toute l'entreprise
- Renseignement en temps réel des tableaux de bord utilisés par la gestion des processus métier (BPM) et les analyses à la demande, garantissant ainsi l'intégration des informations entre les systèmes stratégiques et les applications Web, donc la possibilité pour les employés, clients et partenaires de bénéficier d'un accès aux données en temps réel
- Consolidation des données financières contenues dans divers systèmes répartis entre plusieurs régions, services ou unités commerciales
- Amélioration des performances opérationnelles des systèmes jusque là négativement impactées par la forte réduction des périodes nocturnes de traitement par lots et le recours à des fonctions onéreuses de reporting et de requête

## Accès à toutes les informations

Un grand nombre d'entreprises commencent à considérer leurs entrepôts de données comme une formidable ressource dont le potentiel reste largement sous-exploité. Disposer d'informations justes ne suffit pas : elles doivent être accessibles aux décideurs qui en ont besoin au moment précis où ils en ont besoin ! Or, la plupart des informations nécessaires aux décisions stratégiques sont enfermées dans des entrepôts inaccessibles.

D'après une étude menée par l'agence Gartner Group, moins de 8 % des membres du personnel ont accès aux informations contenues dans l'entrepôt de leur entreprise.

Les applications analytiques ou de BI constituent la première solution de transmission des informations abordée dans le présent guide. Comme mentionné précédemment, les utilisateurs qui interagissent avec de telles applications ont besoin d'accéder à des informations souvent stockées dans des emplacements disparates et qui nécessitent à ce titre d'être intégrées et transformées. L'un des défis restant alors à relever consiste à s'assurer que ces applications ne finissent pas par constituer elles-mêmes un nouveau silo informatique. Les entreprises associent actuellement leurs systèmes opérationnels à leurs systèmes décisionnels, d'où une meilleure exploitation des informations et donc une compétitivité accrue face à la concurrence. Ces entreprises sont capables de comprendre les faits qui motivent leurs décisions avant d'entériner définitivement ces dernières. De nombreux autres processus métier et applications doivent cependant également pouvoir bénéficier d'un accès à des informations fiables. Comment dès lors éviter que chacune de ces applications intègrent, transforment ces données et n'y accèdent de différentes manières ? Comment en effet garantir que les nouvelles voies stratégiques d'acheminement des informations ne deviennent à leur tour incohérentes ? Dans un tel cas de figure, la méthode la plus efficace pour associer ces deux types de systèmes consiste à faire appel à un service Web.

## Flexibilité et réutilisation, deux leitmotiv qui favorisent l'adoption de solutions SOA

- Ressources réutilisables riment avec coûts réduits
- Standardisation rime avec flexibilité
- Combinaison et reconfiguration rapides gages de flexibilité accrues
- Sources transparentes pour une plus grande disponibilité
- Distance vis-à-vis des sources, gage d'une meilleure stabilité
- Ensemble unique de codes, gage d'une moins grande redondance
- Sécurité renforcée et conformité assurée via la non-réitération des informations

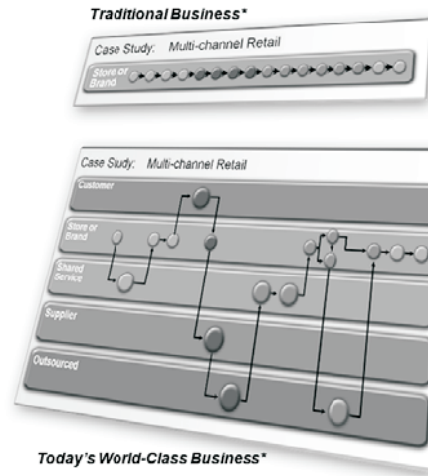


Figure 4.7 : faire appel à un service Web est la meilleure solution dans le domaine des applications opérationnelles

La figure 4.7 présente les avantages qui résultent du déploiement, sous forme d'un service Web, des applications analytiques au sein d'applications opérationnelles. Intégrée à la solution Information Server, l'application IBM InfoSphere Information Services Director permet l'édition des processus d'intégration des données sous forme de services réutilisables au sein d'une architecture SOA (architecture orientée services). En faisant de l'information un service, l'application InfoSphere Information Services Director offre en outre aux entreprises les garanties suivantes : leurs applications et processus métier sont toujours alimentés par des informations métier exactes, cohérentes et fiables, elles peuvent répondre plus rapidement aux nouvelles exigences de leur secteur d'activité, accélérer leurs cycles de développement et réduire leurs dépenses informatiques. L'application InfoSphere Information Services convertit l'intégration des données en un service sans code manuel pouvant être développé en quelques minutes, et ce néanmoins dans un format utilisable par les environnements SOA. Les processus métier et services d'information peuvent ainsi être réutilisés et développés rapidement. Il en découle logiquement pour les entreprises une multitude d'avantages : amélioration de leur flexibilité/réactivité et réduction de leurs cycles de développement, productivité en hausse des équipes informatiques et réduction des coûts afférents au développement.

Grâce à l'application Information Services Director, les entreprises peuvent développer des services Web simples à même de transmettre des faits isolés ou suffisamment d'informations révélant une tendance. Entièrement réutilisables, ces services peuvent servir de pierre angulaire à l'élaboration d'un environnement SOA plus robuste. Ils peuvent également encapsuler des services de fédération, d'intégration des données ou de nombreux autres services. Les architectures SOA décomposent les applications en services, puis permettent à ces services de fonctionner ensemble sous la forme d'applications composites pouvant être construites rapidement en fonction des besoins en constante évolution des entreprises. Dans le cadre de cette approche, les entreprises développent actuellement des services d'information SOA afin d'atteindre les objectifs suivants : rendre leurs employés plus autonomes et plus efficaces et renforcer les processus dynamiques favorisant l'optimisation de leurs activités. Capables de comprendre, nettoyer, transformer et transmettre des données d'entreprise intégrées sous la forme d'un seul service, les services d'information SOA peuvent transmettre directement aux applications des informations fiables.

Or, la transmission de données cohérentes, justes et exhaustives joue un rôle particulièrement important dans le domaine des données de référence, importance qui est d'ailleurs l'objet du prochain chapitre. Ces qualités sont de toute façon indispensables, quelle que soit la nature des données considérées (informations non structurées dans le cadre d'applications de prêt, données financières intégrant des indicateurs de performance clés, etc.). Le défi à relever consiste alors à garantir *un accès cohérent à des informations fiables* tandis que de vastes quantités de données sont disponibles.

## **Automatisation de l'infrastructure informatique**

Partie intégrante de la solution Information Server, l'application DataStage permet de convertir les données quel que soit leur niveau de complexité, et ce à partir de n'importe quelle source, en informations exploitables par les processus décisionnels. La solution IBM Information Server intègre des centaines de fonctions de transformation préélaborées et orientées métadonnées capables d'associer, de restructurer et d'agréger des données clés à partir de leur forme originale dédiée à une application spécifique et ainsi de les adapter à des contextes d'utilisation totalement inédits. Ces informations ainsi traitées peuvent alors être utilisées de manière différente pour une réponse adéquate aux besoins en constante évolution des entreprises. En transformant et standardisant les données de sorte à pouvoir les utiliser dans plusieurs systèmes métier différents, la solution IBM Information Server automatise et simplifie l'intégration des données (plus besoin de les intégrer à partir de sources d'information disparates). Cette solution garantit également un format de données adapté à l'usage envisagé et fournit à tout instant et à tous les utilisateurs autorisés tous les renseignements métier stratégiques dont ils ont besoin.

Grâce à la solution Information Server, les entreprises peuvent obtenir plus facilement des données fiables et intégrées et développer une stratégie de gouvernance facilitant la gestion et l'administration de leurs informations. Traiter l'ensemble des données via un processus unifié de détection, nettoyage et transformation renforce la fiabilité et la précision des données. Fournir aux décideurs des informations dans lesquelles ils peuvent avoir une confiance totale ouvre la voie à des décisions sectorielles plus audacieuses et plus innovantes.

### **Déploiement et gestion unifiés des métadonnées**

L'intégration des données est un processus qui nécessite d'être élaboré à partir d'une infrastructure unifiée des métadonnées où le rôle des différents utilisateurs impliqués dans le processus doit être clairement compris, quel que soit le domaine concerné, métier, opérationnel, technique, etc.

Cette infrastructure gérée et partagée favorise la réduction des délais de développement et offre un fichier de suivi permanent capable de renforcer la confiance dans les données disponibles tout en limitant les opérations manuelles fastidieuses de coordination et de synchronisation à effectuer.

La coordination des services métier et informatiques était, rappelons-le, l'un des principaux défis à relever mentionné en préambule du présent chapitre. Divers groupes d'utilisateurs dont les rôles varient grandement, que ce soit du côté métier ou du côté informatique, doivent en effet pouvoir échanger sans heurts des informations directement les uns avec les autres, et ce par-delà les frontières organisationnelles et fonctionnelles. Une collaboration optimisée entre un grand nombre d'utilisateurs différents est d'ailleurs essentielle à la réussite de tout projet d'intégration. Tous les modules de la solution Information Server



utilisent l'application Metadata Workbench (fig. 4.8). Celle-ci permet de conserver les informations issues de chaque étape et donc de les réutiliser ultérieurement tout en offrant de précieux enregistrements aux fins de traçabilité et de vérifiabilité.

## InfoSphere Information Server



Figure 4.8 : solution InfoSphere Information Server avec application Metadata Workbench

## Regroupement des données aux fins d'analyse

Regrouper des données historiques aux fins d'analyse correspond au processus baptisé *entrepotage de données*. Les entreprises s'attachent actuellement à moderniser leurs infrastructures et pratiques commerciales afin de modifier radicalement la manière dont elles font des affaires. Les stratégies de modernisation adoptées visent à améliorer sensiblement leur réactivité/flexibilité et efficacité tout en réduisant leurs coûts, en d'autres termes à accomplir davantage avec moins de ressources, ainsi qu'à modifier les processus métier et normes du marché. Les anciens systèmes décisionnels se limitent à des fonctions de reporting et offrent peu de solutions en matière d'analyse des données prospectives et d'analyse des processus en cours. En outre, les résultats des quelques analyses effectuées sont généralement uniquement accessibles à moins de 8 % des employés de l'entreprise.

En acteurs judicieux et éclairés, les nouvelles entreprises mettent en œuvre des méthodes innovantes de diffusion de l'information au sein de tous leurs services qui leur permettent de prendre en toute confiance et en toute connaissance de cause des décisions avisées. Pour résoudre les problèmes auxquels elles sont actuellement confrontées, elles font appel à des données et informations à jour. Les détaillants utilisent les informations dont ils disposent pour vendre à leurs clients d'autres produits susceptibles de les intéresser au

moment de leur achat initial. Les compagnies d'assurance détectent et préviennent les fraudes au niveau du point de contact avec leurs assurés. Les policiers identifient les suspects éventuels dès leur arrivée sur les scènes de crime. Les fabricants limitent leurs stocks en fonction du nombre de commandes prévu. Ces capacités sont autant d'atouts à même de bouleverser les règles du jeu, à condition cependant qu'elles reposent sur des informations fiables.

## ***Appliances et entreposage des données d'entreprise***

Collecter des informations afin d'en dériver de la valeur ajoutée est l'objectif de l'entreposage des données depuis de nombreuses années. Le problème d'un tel processus réside dans le fait qu'il est long, onéreux et complexe. Les appliances d'entreposage des données doivent leur popularité aux fonctionnalités de déploiement plus rapides qu'elles proposent, d'où la possibilité pour les entreprises de récolter les avantages en termes de valeur ajoutée plus tôt. Offrant la simplicité d'une solution intégrée, ces appliances permettent également de réduire les coûts d'exploitation des entreprises, la gestion et la maintenance du système nécessitant moins de personnel et d'efforts. Ces dispositifs, qui restent un sujet sensible sur le marché, sont une solution qui dépasse le simple cadre des magasins de données puisqu'ils sont intégrés au système même d'entreposage des données d'entreprise. Les entrepôts de données mis en œuvre à l'heure actuelle doivent désormais être reliés aux systèmes opérationnels et réciproquement, de sorte que les opérations quotidiennes des entreprises soient alimentées en retour par les analyses issues de ces entrepôts. On parle alors parfois de *BI opérationnelle*, approche considérée par IBM comme l'un des éléments clés de l'entreposage dynamique des données. Cette intégration entre systèmes opérationnels d'un côté et systèmes d'entreposage de l'autre nécessite le recours à des technologies plus avancées, c'est-à-dire à des technologies capables d'offrir des mises jour et analyses instantanées, gages d'une optimisation des processus décisionnels. Les analyses intégrées au sein des entrepôts de données font en outre l'objet d'une attention croissante. En raison de l'entreposage accéléré des données, leur extraction vers des outils distants et le rechargement des résultats obtenus ne sont plus possibles car ces deux processus prennent trop de temps : en un mot, les décideurs souhaitent obtenir la réponse à leurs problèmes quotidiens, le jour même de leur survenue. Autre tendance qui se dessine à l'horizon : le besoin croissant d'effectuer des analyses sur des données non structurées. Ce type de données peut constituer jusqu'à plus de 80 % des informations dont disposent les entreprises et peut ne pas être compatible avec les méthodes d'entreposage habituelles. On comprend dès lors aisément pourquoi un grand nombre d'entreprises leaders sur le marché se tournent vers des solutions d'entreposage des données stratégiques pour les aider à construire l'entreprise intelligente de demain.

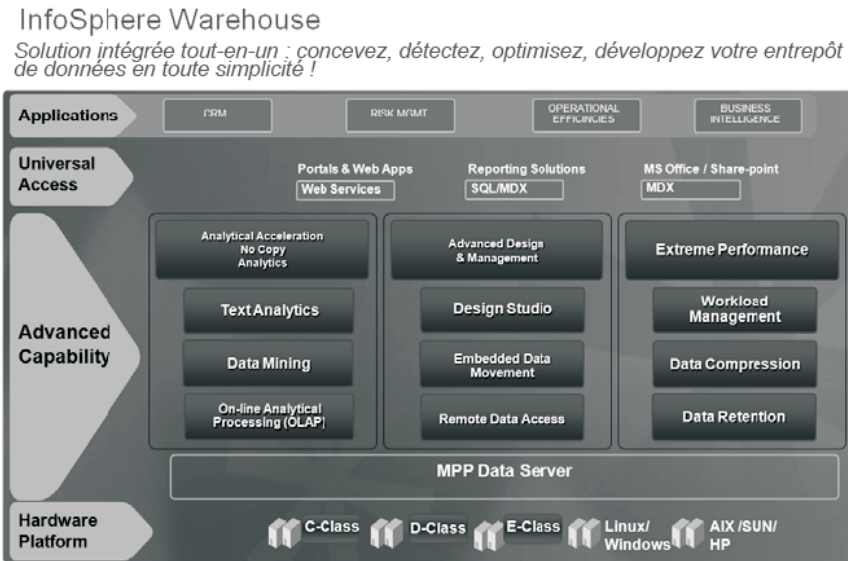


Figure 4.9 : InfoSphere Warehouse

## InfoSphere Warehouse au service d'une entreprise intelligente

La solution IBM InfoSphere Warehouse traite de manière exhaustive tous les problèmes posés par la gestion des informations et abordés tour à tour au cours des chapitres précédents (fig. 4.9). Environnement multi-usage, complet, flexible et robuste, la solution InfoSphere Warehouse offre à l'entreprise intelligente de demain la possibilité d'accéder à des informations ultra actualisées, d'analyser ces types d'informations et d'interagir avec celles-ci, et ce quel que soit leur type : structuré, non structuré, opérationnel, transactionnel, actuel ou historique. Grâce à la solution InfoSphere Warehouse, les entreprises peuvent analyser et optimiser l'intégralité de leurs processus métier de gestion de la performance. Cette situation profite également largement aux utilisateurs de l'entreprise puisque, grâce à InfoSphere Warehouse, ils bénéficient d'une solution à la pointe de l'innovation dans le domaine de la planification, de la budgétisation, des prévisions, de la modélisation, de la consolidation et du reporting financier. Autre avantage majeur, l'intégration d'informations issues de différentes sources donne la possibilité aux entreprises d'identifier les segments du marché, produits et canaux de distribution les plus rentables. Ces données intégrées permettent également d'explorer de nouvelles opportunités commerciales dans les domaines suivants : ventes croisées, compréhension anticipée des nouveaux besoins émergents sur le marché et développement de programmes marketing offrant des produits et services en phase avec

les attentes des clients. La solution IBM InfoSphere Warehouse peut être exécutée sur des plates-formes distribuées comme dans des environnements mainframe.

## **Gestion du cycle de vie des informations contenues dans les entrepôts de données**

*La gestion du cycle de vie des informations* correspond pour l'essentiel à la capacité à surveiller, analyser et gérer le cycle de vie des applications, systèmes et processus s'exécutant dans l'environnement d'entreposage. Cette capacité à contrôler et optimiser l'environnement d'entreposage comprend également la définition de règles à même d'optimiser l'utilisation des ressources système disponibles (UC, mémoire, disque). La mise en œuvre d'un système global de gestion du cycle de vie au sein des entrepôts de données offre aux entreprises le double avantage suivant : réduction des coûts et système capable de supporter un environnement d'entreposage dynamique hautes performances.

Toute gestion efficace du cycle de vie des informations des entrepôts intègre nécessairement les éléments suivants :

- Optimisation du stockage :
  - » Limitation optimale des besoins en termes de stockage et de mémoire sans détérioration des performances
- Gestion de la performance :
  - » Surveillance et analyse de l'environnement en vue d'optimiser la performance
- Rétention des données :
  - » Gestion du cycle de vie des données basée sur des règles en vue d'une réduction des frais de stockage et d'extraction
- Gestion de la charge de travail :
  - » Gestion et contrôle des ressources informatiques pour une réponse adaptée aux diverses requêtes émises par les charges de travail

## **Master Data Management (gestion de données de référence)**

Les données de référence sont des informations métier clés, par exemple des données client et/ou des données produit, généralement réparties entre plusieurs applications et sites. Utilisées habituellement par plusieurs applications appartenant à divers secteurs d'activité de l'entreprise, ces informations sont généralement gérées service par service et non à l'échelle de l'entreprise dans son ensemble (par exemple, les clients d'une même banque peuvent être gérés simultanément par différents services : prêts, épargnes, cartes de crédit, etc.). Cette

gestion multiservice peut présenter d'importants inconvénients : données présentes en double, informations inexactes, sans parler de la frustration des clients, fournisseurs et autres partenaires.

Si générer une vue intégrée des données de référence contenues dans l'entrepôt permet d'obtenir une version unique de la réalité, cette solution offre peu d'avantages sur le plan des applications opérationnelles. Charger ces données intégrées dans chaque système opérationnel peut être une possibilité, mais chacun de ces systèmes obéissant à ses propres règles métier, les données ainsi chargées risquent rapidement de ne plus être synchronisées. Le recours à un système capable de gérer en toute sécurité les informations pour l'ensemble des applications est en fait la solution à ce problème.

#### Avantages de la solution Master Data Management

- Dissocie les données de référence des applications individuelles
- Devient une ressource centrale, indépendante vis-à-vis des applications
- Simplifie les tâches d'intégration en cours et le développement de nouvelles applications
- Garantit la disponibilité de données de référence cohérentes entre les systèmes consacrés aux transactions et aux analyses

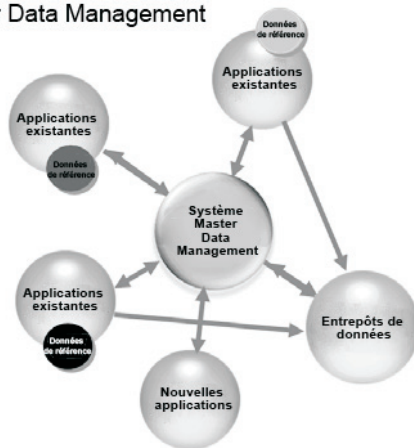


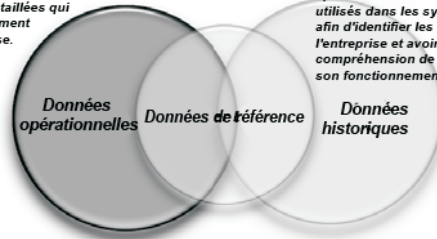
Figure 4.10 : Master Data Management

Modèle bien établi sur le marché, la solution Master Data Management (fig. 4.10) offre un référentiel pour le stockage des données ainsi qu'une grande bibliothèque de services Web SOA permettant la lecture, l'insertion, la mise à jour et la gestion de chacun des champs utilisés pour les données de référence.

Les données de référence sont présentes aussi bien dans les entrepôts de données que dans les systèmes opérationnels. A titre d'exemple, lorsque M. Jean Martin se rend chez un détaillant pour y effectuer un achat, les informations le concernant directement correspondent aux données de référence tandis que les informations concernant son achat correspondent aux données opérationnelles. Une fois l'achat effectué, ces dernières informations deviennent des données historiques et sont transférées jusqu'à l'entrepôt des données (fig. 4.11). En résumé, les données relatives à l'achat sont d'abord opérationnelles avant de devenir historiques tandis que les informations relatives à M. Jean Martin conservent leur statut de données de référence.

## Comparaison entre les données de référence et les données opérationnelles et historiques

Les données opérationnelles sont des données métier détaillées qui assurent le fonctionnement quotidien de l'entreprise.



Les données historiques correspondent aux artefacts des données opérationnelles. Ces artefacts sont utilisés dans les systèmes décisionnels afin d'identifier les tendances de l'entreprise et avoir une meilleure compréhension de ses activités et de son fonctionnement.

Les données de référence sont des informations métier clés, par exemple des données client et/ou des données produit, généralement réparties entre plusieurs applications et sites.

Figure 4.11 : comparaison entre les données de référence et les données opérationnelles et historiques

Les solutions MDM (Master Data Management) se développent à partir des bases jetées pour l'intégration des données et établissent les critères de fiabilité des informations ou définissent les principales entités de données de référence de l'entreprise, garantissant ainsi la cohérence des entités métier les plus importantes au niveau de toutes les applications de gestion de la performance et à l'échelle de toute l'entreprise.

Les principales raisons qui poussent les entreprises à adopter une solution MDM sont les suivantes :

- Réduction des coûts
- Baisse des coûts
- Relations, compréhension et rétention améliorées
- Hausse des recettes
- Initiatives stratégiques facilitées

Solutions complémentaires, l'entreposage des données et les systèmes MDM développent des données d'entreprise exploitables aux fins d'analyse et de reporting. Les solutions MDM assurent la collecte et la gestion des informations utilisées par les principales entités métier (clients, produits, sites, etc.). Associées aux données historiques des entrepôts de données, les données de référence permettent d'avoir une compréhension approfondie des comportements et préférences des clients. Les systèmes MDM sont également utiles aux applications et systèmes opérationnels lorsque ceux-ci souhaitent obtenir un aperçu opérationnel unique de leurs données clés. Aperçu unique grâce auquel

les entreprises peuvent ensuite vérifier que les informations utilisées lors d'opérations de routine telles que l'ouverture d'un compte sont cohérentes et exactes.

Obtenir des informations cohérentes, exactes et exhaustives est une tâche particulièrement ardue pour les entreprises, nombreuses, qui ne disposent pas d'une vue d'ensemble de leur organisation ni de leurs besoins en termes d'information. Les données de référence sont par définition des données omniprésentes, c'est-à-dire des données présentes dans pratiquement tous les systèmes. Souvent isolés les uns des autres et non pourvus de fonctions appropriées de contrôle de l'intégrité des informations, ces systèmes contiennent souvent des données de référence incohérentes, inexactes et incomplètes. Un grand nombre d'entreprises commencent à s'intéresser de plus en plus à ces données spécifiques tandis qu'elles essaient de se conformer aux nouvelles réglementations en vigueur ou souhaitent effectuer des opérations de fusion. En l'absence de système MDM, obtenir un aperçu opérationnel unique des clients, produits ou sites relève de l'exploit.

Quel que soit l'objectif final du système de données de référence envisagé ; applications d'analyse, systèmes de reporting ou systèmes opérationnels, le dénominateur commun indispensable est toujours le même : les utilisateurs doivent avoir pleinement confiance dans les données dont ils se servent. La possibilité de disposer d'informations compréhensibles, exactes, opportunes et contextualisées est le moteur clé de la réussite de tous les leaders du marché. Divers outils et approches sont possibles en matière d'extraction, de transformation et de chargement des données : quelle que soit la solution adoptée, le point essentiel à ne pas perdre de vue est le fait suivant : les utilisateurs doivent en toutes circonstances avoir confiance dans les informations disponibles et ainsi pouvoir prendre des décisions audacieuses à même de provoquer des changements favorables au sein de l'entreprise : réduction des coûts, identification anticipée des nouvelles opportunités commerciales et modification des règles du jeu avant que la concurrence n'ait eu le temps d'identifier ces opportunités.

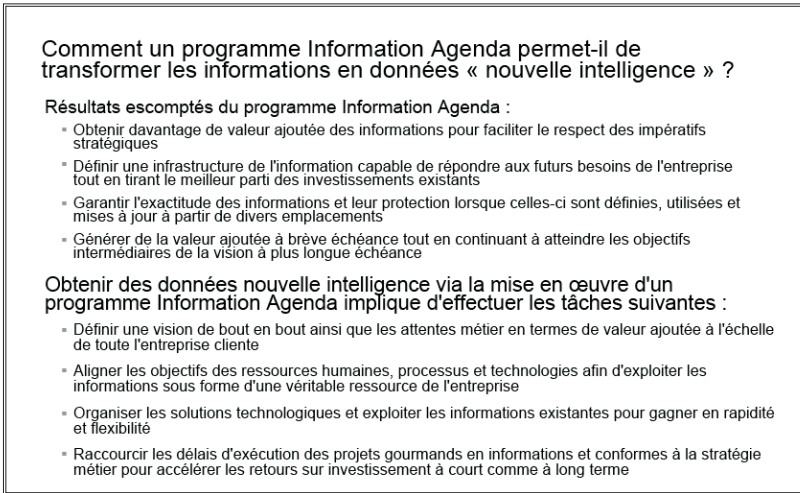
# 5

## Le programme Information Agenda

**D**ans le contexte économique actuel à la fois difficile, dynamique et mondialisé, les entreprises reconnaissent de plus en plus que pour rester compétitives elles doivent faire un meilleur usage des informations à leur disposition. Cependant, si la plupart des entreprises ont déjà mis en place des stratégies de gestion pour leurs processus et applications métier, les informations requises en renfort sont elles rarement partagées entre toutes les entités. En outre, l'absence d'une stratégie de gestion cohésive empêche d'exploiter pleinement les données disponibles.

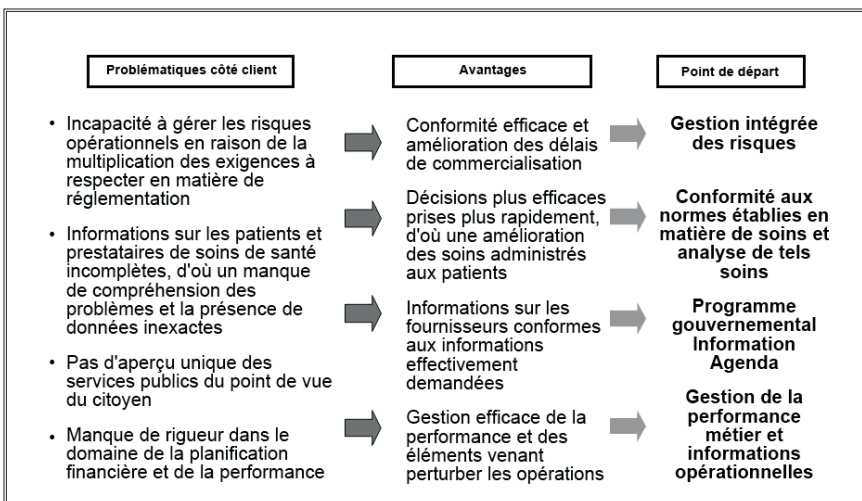
Approche éprouvée, le programme de gestion de l'information Information Agenda aide les entreprises à transformer leurs informations en données « nouvelle intelligence ». Gage d'objectifs plus facilement atteints et de performances améliorées, l'approche plus unifiée de ce programme en matière d'exploitation des informations rend également les entreprises plus perspicaces (fig. 5.1).





*Figure 5.1 : nouvelle intelligence générée grâce au programme Information Agenda*

La figure 5.2 illustre comment un programme Information Agenda peut être appliqué à différents secteurs d'activité afin de favoriser le lancement de nouveaux projets et permettre à l'entreprise de bénéficier de nouveaux avantages.



*Figure 5.2 : les entreprises complexes et les entreprises présentant des silos informatiques doivent faire face à des problèmes spécifiques que les solutions Information Agenda/nouvelle intelligence peuvent résoudre*

La Figure 5.3 illustre les composants d'un programme Information Agenda à même de fournir des informations d'entreprise fiables et dynamiques capables de favoriser et soutenir le développement d'applications nouvelle intelligence.

**Approche éprouvée :**

*Un savoir-faire et des compétences garantissant la réussite et la réduction des risques*

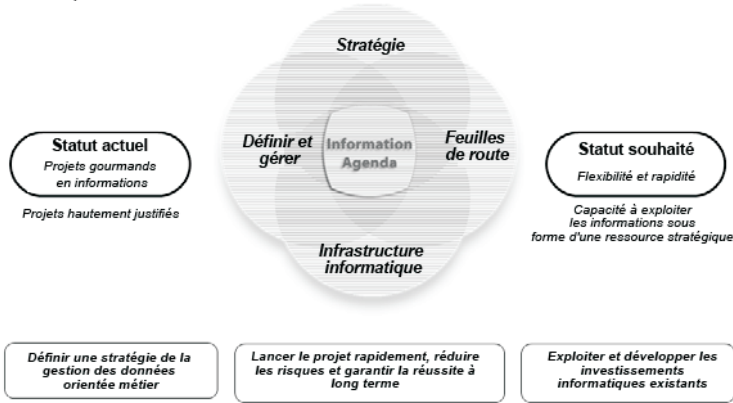


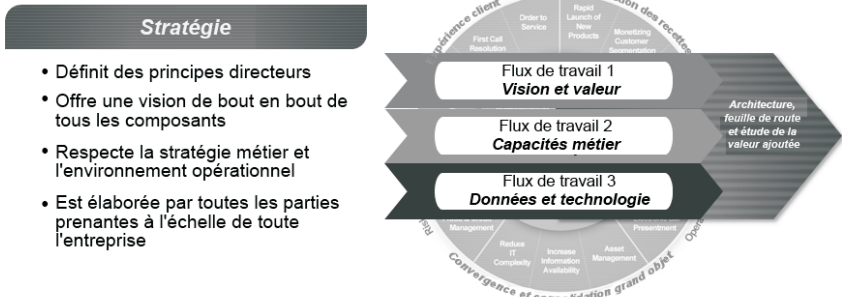
Figure 5.3 : le cadre de travail du programme Information Agenda

## Stratégie de gestion des données

Les principes de la stratégie de gestion des données adoptée (fig. 5.4) permettent d'orienter les efforts, en matière de création et d'utilisation des données, consentis par l'entreprise. Ces efforts ont également pour but de soutenir les projets, objectifs et initiatives de cette dernière. Reposant sur la stratégie métier et l'environnement opérationnel de l'entreprise, la stratégie de gestion des données offre une vision de bout en bout de tous les éléments composant le programme Information Agenda.

Tandis qu'elle définit cette stratégie spécifique, l'entreprise doit en profiter pour identifier les objectifs et initiatives sources d'innovation, capables de faire la différence sur le marché. Ces objectifs et initiatives correspondent en fait à l'orientation, dans les grandes lignes, que l'entreprise, d'après les dirigeants, doit prendre pour exploiter au maximum ses forces et opportunités inhérentes face à la concurrence. La stratégie de gestion des données doit être régulièrement réajustée en fonction de l'évolution de la stratégie métier adoptée par l'entreprise.

## Une stratégie de gestion des données offre une vision globale et assure le lien avec la stratégie métier



La collaboration entre le DSI et les responsables des autres branches d'activité de l'entreprise est un facteur clé essentiel à la réussite de l'Information Agenda (programme de gestion de l'information), toutes entreprises confondues.

Figure 5.4 : objectif de la stratégie de gestion des données

## Feuilles de route

Les feuilles de route de gestion de l'information (fig. 5.5) permettent d'identifier les suites récurrentes de projets et d'objectifs produisant des résultats tangibles. Elles peuvent également permettre au DSI d'identifier efficacement les attentes ainsi que les secteurs nécessitant un financement. Elles doivent enfin prendre en compte les opportunités à courte comme à longue échéance et fournir un processus d'évaluation et de suivi des avantages métier.

Exemple de feuille de route d'un programme Information Agenda  
 Exploitation de l'intelligence métier / Définition des bases pour la gestion de la performance et nouvelles initiatives orientées vers l'avenir

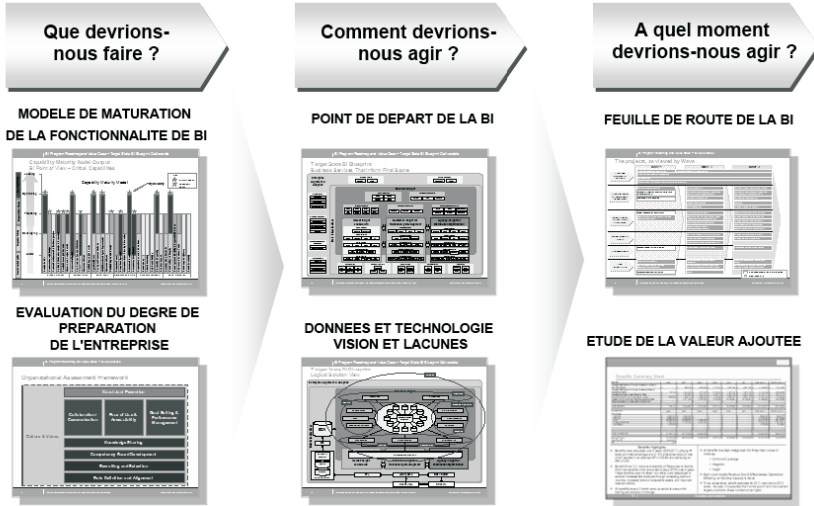


Figure 5.5 : feuilles de route de gestion de l'information

## Infrastructure informatique

Pour faire de l'information une ressource stratégique et parvenir à gérer cette ressource dans la durée, les entreprises doivent mettre en œuvre une infrastructure informatique exploitable à grande échelle (fig. 5.6). L'adoption d'architectures limitées prenant uniquement en compte des applications spécifiques aboutit généralement à d'importants dysfonctionnements opérationnels, à des délais d'exécution inacceptables pour les projets ainsi qu'à la multiplication incontrôlée du nombre de copies d'un même contenu. Dans le cadre d'un programme Information Agenda, les infrastructures informatiques identifient en revanche les technologies nécessaires à l'intégration des investissements actuels aux futures technologies, d'où la garantie d'un retour sur investissement optimisé.

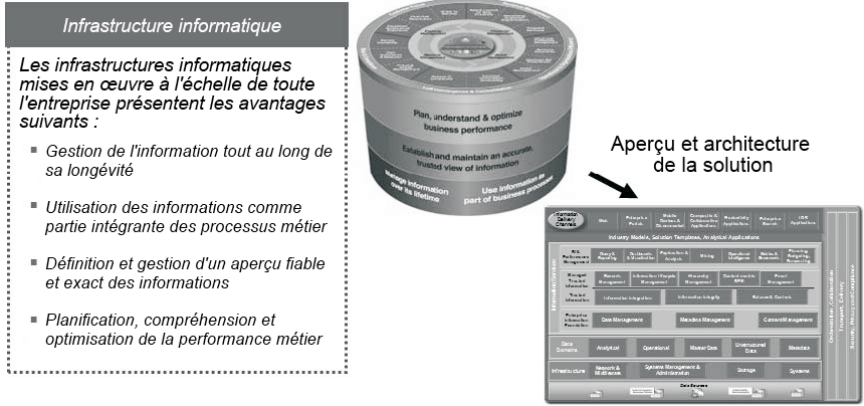


Figure 5.6 : la mise en œuvre d'une infrastructure informatique à l'échelle de toute l'entreprise est un facteur clé de réussite pour assurer la gestion efficace des informations dans la durée sous forme de ressource stratégique

## Gouvernance informatique

Une fois les objectifs coordonnés, les feuilles de route définies et l'infrastructure informatique mise en place à l'échelle de toute l'entreprise, la gestion adoptée joue un rôle clé dans l'utilisation du programme Information Agenda lorsqu'il s'agit de préserver la valeur des données au fil du temps. A l'instar de toutes les autres ressources de l'entreprise, les informations doivent faire l'objet d'une gestion attentive. Ce type de gestion, baptisé *gouvernance informatique*, améliore la qualité, la disponibilité et l'intégrité des informations disponibles au sein des entreprises en favorisant la définition concertée de politiques et la coopération entre services (fig. 5.7).

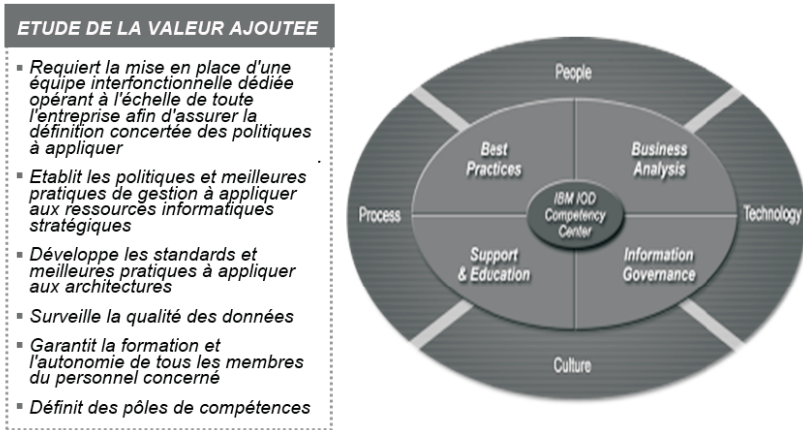


Figure 5.7 : une gouvernance informatique efficace permet d'améliorer la qualité, la disponibilité et l'intégrité des informations disponibles au sein des entreprises

Bien que la coopération entre les DSI et les responsables des branches d'activité afin d'établir les standards de la gouvernance informatique ne soit pas toujours facile à obtenir, son importance dans le cadre du programme Information Agenda ne doit pas être sous-estimée. La mise en œuvre d'une gouvernance informatique robuste est indispensable pour assurer une mise en conformité plus rapide et plus globale des entreprises aux réglementations externes en vigueur. Ce type de gouvernance permet également de récolter enfin les avantages financiers résultant d'une plus grande qualité des données, d'une amélioration des processus de gestion et d'une responsabilité accrue. La mise en place, toujours dans le cadre de la gouvernance informatique, de définitions et processus partagés favorise également une plus grande efficacité au niveau de l'élaboration, de l'exécution, du suivi et de la gestion des stratégies, d'où une amélioration de la performance globale de l'entreprise.

## Concrétisation des objectifs

Les DSI qui souhaitent, pour rendre les données plus fiables, développer et mettre en œuvre un programme Information Agenda efficace, en phase avec les objectifs métier de l'entreprise, ont besoin de l'appui des responsables des divisions métier. Ils doivent donc obtenir le soutien d'un dirigeant haut placé, par exemple d'un membre du comité de direction, PDG, directeur financier ou autre capable de faire du programme Information Agenda une initiative stratégique de l'entreprise. Grâce à cette adhésion au plus haut niveau, les DSI sont assurés d'obtenir le soutien des responsables des autres secteurs d'activité de l'entreprise et de toute personne dont la participation et l'adhésion au développement sont essentielles à la réussite du programme.

Bénéficiant du soutien non seulement de la direction mais également de celui d'une équipe de collaborateurs motivés, les DSI peuvent ainsi accélérer la mise en œuvre du programme Information Agenda. Pour développer un tel plan de mise en œuvre, différentes étapes doivent être respectées :

1. Vision partagée et soutien de la direction
2. Analyse de la stratégie d'entreprise et des facteurs de progression
3. Evaluation de référence (capacités et maturation)
4. Développement de l'état futur, identification des opportunités et hiérarchisation des priorités
5. Elaboration intégrée des feuilles de route et planification de la mise en œuvre
6. Coup d'envoi du projet

Ces étapes offrent une approche structurée, modulaire et accélérée du programme Information Agenda lorsqu'il s'agit de définir, d'améliorer, de justifier et d'exécuter ce programme. Elles sont spécifiquement conçues, d'une part, pour réconcilier les divers aspects informatiques et métier impactant le programme Information Agenda et, d'autre part, pour produire des résultats spécifiques, notamment les résultats suivants :

- Etudes de cas sectorielles
- Evaluation de la valeur ajoutée métier
- Evaluation du degré de préparation de l'architecture existante Solution architecture
- Choix de l'architecture
- Feuille de route intégrée pour le projet
- Services et ressources logicielles

## **Conclusion**

L'environnement dans lequel les entreprises doivent opérer aujourd'hui est radicalement différent de ce qu'il était auparavant : nouvelles règles économiques, internationalisation, éléments fortement liés les uns aux autres, risques accrus et explosion des volumes de données disponibles. Les dirigeants qui doivent prendre chaque jour des décisions

importantes ont conscience que certaines zones d'ombre subsistent lorsqu'ils prennent ces décisions alors même que la moindre petite erreur peut être gravement sanctionnée, que les coûts doivent être éradiqués des systèmes et les décisions prises à un rythme de plus en plus échevelé. L'intuition et l'instinct, la plupart du temps, fruits de l'expérience personnelle, ne suffisent plus à une prise de décision pertinente.

L'avènement d'une nouvelle entreprise plus perspicace, c'est-à-dire d'une entreprise « nouvelle intelligence » se servant plus judicieusement des faits disponibles est essentiel et possible, grâce à l'exécution à grande échelle d'analyses avancées sur des jeux de données largement plus riches et plus intégrés. Les entreprises peuvent faire de cet idéal une réalité. Comment ? Tout simplement en associant à la gestion de la performance et principales données métier les éléments suivants : analyses avancées, modélisation prédictive, solutions logicielles d'envergure internationale et modèles éprouvés diminuant les délais de valorisation pour les clients.

Les dirigeants doivent revoir la portée traditionnelle de leurs activités afin que celles-ci intègrent désormais les changements intervenus sur la définition des concepts suivants : information, distance, vitesse, collaboration, contrôle et accès. Pour exister dans un monde intelligent, ils doivent inventer des centres de données décloisonnés, être prêts à gérer de nouvelles sources d'information inconnues, identifier les points d'entrée des composants intelligents et réfléchir à l'image qu'ils diffusent de leur entreprise en termes de performance et d'attitude citoyenne. Les progrès réalisés en matière d'intégration des systèmes, de développement de réseaux et de sécurité vont donner naissance à des infrastructures rapides et pertinentes plus intelligentes.



# Une nouvelle intelligence pour une planète plus intelligente

## Les solutions analytiques IBM comme moteurs de l'innovation stratégique

Tout responsable prend des décisions. Et toute décision dépend des informations. Ceci se vérifie à chaque fois que quelqu'un dirige une entreprise, un gouvernement ou un foyer.

Au cours des 50 dernières années, les responsables ont vu une révolution de la qualité et la quantité des informations à leur disposition. L'ère industrielle a été supplantée par l'ère de l'information ; ceci est dû en grande partie au fait que la concurrence sévit désormais au niveau des connaissances, de l'expertise et du capital intellectuel. En réalité, la valeur ultime du secteur de la technologie de l'information n'a jamais concerné les puces, les ordinateurs ni les logiciels. Le secteur a toujours cherché à aider les responsables à connaître, en toute confiance, les événements passés, présents et à venir, et ce quel que soit l'aspect de l'entreprise concerné. Voyez-vous les schémas clés ? Pouvez-vous extraire des idées essentielles à partir des données ? Pouvez-vous prendre un peu de recul et calculer le coût de la prise et de l'application d'une décision ?

Quelle que soit la mesure, ce type de questions trouve des réponses bien plus précises qu'auparavant. Le défi évolue au fur et à mesure que le volume et la variété des informations croissent, ainsi que la vitesse de la prise de décisions sur une planète plus intelligente.

Heureusement, la technologie existe pour capturer et traiter toutes ces données, puis les transformer non seulement en informations organisées, ou même en connaissances, mais en intelligence réelle. Nous pouvons alors distinguer des schémas de façon bien plus détaillée qu'auparavant. Nous pouvons capturer et analyser les évolutions des marchés, des tendances et des préférences des consommateurs bien plus rapidement qu'avant. Cette nouvelle intelligence a la possibilité de changer le monde et la façon dont il tourne.

**PAT BATES** Spécialiste du groupe  
Worldwide Technical Sales Enablement

**MIKE BIERE** Directeur marketing,  
des données et Business Intelligence  
pour System z

**REX WIEDERANDERS** Directeur du groupe Worldwide  
Technical Sales Enablement

**ALAN MEYER** Directeur marketing,  
Entreposage de données InfoSphere

**BILL WONG** Cadre dirigeant,  
Gestion des informations, solutions  
de transformation induite par l'information



MC Press Online, LP  
125 N. Woodland Trail  
Lewisville, TX 75077

Price: \$14.95 US/\$16.95 CN



IMM14055-FRRF-00