
Livre blanc : la business intelligence

Charlotte GALZY

Pierre GIRONA

Benoît MARTIN

Cédric NICOLOSO

Julien VANDERMOERE

Mai 2010

Sommaire

Introduction	5
Les ETL.....	6
Introduction	6
Marché des ETL.....	7
Les perspectives et les challenges de l'ETL	8
Les nouveaux modes de stockage.....	9
Introduction	9
Les acteurs	9
Qliktech.....	9
SAP	9
La technologie « In-memory »	10
Les solutions.....	10
Qlikview.....	10
La solution de SAP.....	11
Conclusion.....	11
La business intelligence en mode SaaS.....	12
Introduction	12
Marché et acteurs.....	13
Les éditeurs et leurs solutions	13
Les cibles	13
Les changements apportés par le Saas	14
Pour les éditeurs/intégrateurs.....	14
Pour les clients	14
Stockage de données et sécurité	17
Conclusion.....	18
La business intelligence mobile	19
Introduction	19
Marché et acteurs.....	20
Marché des smartphones	22
Sécurité	24
Apports de la mobilité.....	25
Conclusion.....	26
Les moteurs de recherche	27
Introduction	27
L'importance du moteur de recherche.....	28

Le marché et les acteurs	29
Les technologies du web sémantique.....	31
Le futur des moteurs de recherche	32
Les médias de restitution.....	33
Introduction	33
Pourquoi l'indexation ?	33
Les acteurs	34
Les solutions.....	35
IMEDIA de l'INRIA	35
Google Audio Indexing (Gaudi)	35
Aigle de Business Geographic	36
Exalead CloudView.....	36
Quaero	37
LTU	37
Chromatic.....	38
Autonomy	38
NewPhenix	39
Google Location Services	39
Conclusion.....	39
Master Data Management.....	40
Introduction	40
Pourquoi le MDM ?	40
Définitions, approches et concepts	42
Normes et standards.....	45
Le marché du MDM	46
Les acteurs	48
Le projet MDM	51
Perspectives et tendances	52
Conclusion.....	54
Sources.....	55
Les ETL.....	55
Les nouveaux modes de stockage.....	55
In memory	55
Qlikview.....	55
SAP	55
La business intelligence en mode SaaS.....	55
Marché et acteurs	55

Changements apportés par le Saas.....	56
Stockage de données et sécurité	56
La business intelligence mobile	56
Acteurs	56
Marché des smartphones	56
Sécurité	56
Apports de la mobilité.....	56
Les moteurs de recherche.....	57
Introduction	57
L'importance du moteur de recherche.....	57
Le marché et les acteurs	57
Les technologies du web sémantique.....	57
Le futur des moteurs de recherche	57
Les médias de restitution.....	57
Indexation	57
IMEDIA	57
CQBI	57
Gaudi.....	57
Aigle	58
Exalead.....	58
Quaero	58
LTU	58
Chromatic.....	58
Autonomy	58
Google Location Services	58
NewPhenix	58
MDM	58
Site d'informations et d'actualités sur la Business Intelligence.....	58
Site d'actualités générale et business.....	58
Site d'éditeur ou de cabinet de conseil	59
Blog business et MDM	59
Livre blanc et livre	59

Introduction

Selon le cabinet Gartner, le marché mondial de la BI¹ s'est élevé à 9,3 milliards de dollars dans le monde en 2009, ce qui, malgré la crise, note une progression de 4,2% par rapport à 2008. Des chiffres relativement réconfortants et qui confirment un marché en plein essor.

Les avancées technologiques ainsi que les exigences du marché ont rendu ce domaine incontournable chez les analystes, les décideurs et les développeurs.

Tout le monde se met à la BI. La valeur ajoutée que peut apporter un environnement de BI dans une entreprise est stupéfiante. Prenons l'exemple de l'entreprise Wal-Mart, une grosse entreprise de distribution grand public. Elle a prouvé cela en démarrant un projet de BI très peu après la création de l'entreprise. Résultat, l'entreprise est indétrônable, sa gestion interne est parfaite, la gestion des stocks impeccable, et sa capacité à prévoir les tendances et habitudes des consommateurs nous met à la fois en admiration et en crainte : ils sont capables de prévoir ce que le client moyen va acheter avant même qu'il ne rentre dans un magasin...

C'est pourquoi nous allons dans ce livre blanc faire une étude plus approfondie de la business intelligence en suivant différents axes d'étude tel que :

- Les ETL ;
- Les nouveaux modes de stockage ;
- La BI en SaaS² ;
- La BI mobile ;
- Les médias de restitution ;
- Les moteurs de recherche ;
- Le MDM³.

¹ Business Intelligence

² Software as a Service

³ Master Data Management

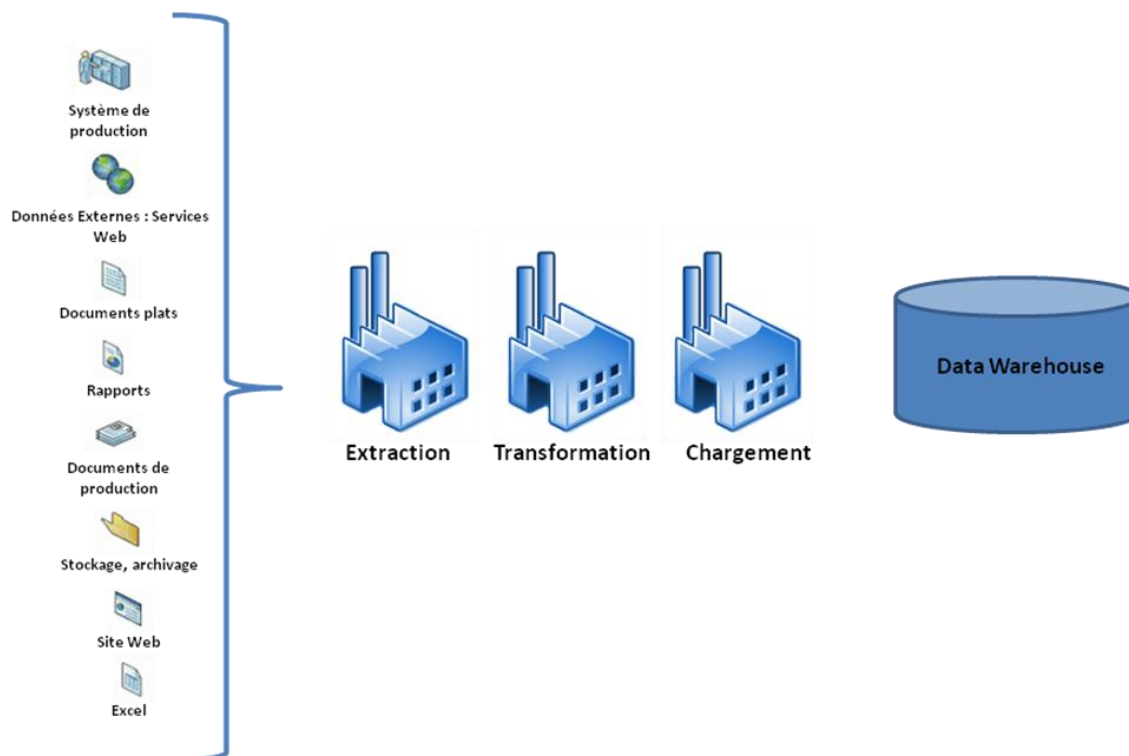
Les ETL

Introduction

Un ETL, acronyme d'Extraction-Transformation-Loading est un système de chargement de données depuis les différentes sources de l'entreprise jusqu'à l'entrepôt de données. Afin de charger correctement ces données, ce système s'occupe de réaliser différentes opérations.

- **EXTRAIRE** : accéder à la majorité des systèmes de stockage de données (SGBD, ERP, fichiers à plat...) afin de récupérer les données identifiées et sélectionnées.
- **TRANSFORMER** : toutes les données ne sont pas utilisables telles quelles. Elles doivent être vérifiées, reformatées, nettoyées et enfin consolidées.
- **CHARGER** : insérer les données dans le datawarehouse. Elles sont ensuite disponibles pour les différents outils d'analyse et de présentation comme les tableaux de bord, l'analyse multidimensionnelle OLAP, le data mining, les requêteurs et autres reportings.

Les termes Back Room (la cuisine) ou Data staging area sont souvent utilisés par l'industrie pour décrire les systèmes ETL.



Les processus ETL sont les composants les plus critiques et les plus importants d'une infrastructure décisionnelle. Bien que cachés de l'utilisateur de la plate-forme décisionnelle, les processus ETL rassemblent les données à partir des systèmes opérationnels et les prétraitent pour les outils d'analyse et de reporting. La précision et la vitesse de la plate-forme décisionnelle toute entière dépendent des processus ETL. D'ailleurs Ralph Kimball, un informaticien et chef d'entreprise américain, après dix huit mois d'études des ETL, en a défini trente-huit sous-systèmes et il a statué que 70% d'un projet d'entrepôt de données est dédié aux systèmes ETL.

Marché des ETL

Aujourd'hui, l'Open Source a pris une part importante du marché et beaucoup d'ETL se fondent au sein d'offres plus globales d'intégration de données.

Synthèse des solutions à suivre

Editeur	Solution	Tarification	Principaux clients
IBM	Datastage Server	Dès 21 000 €	NC
	Datastage PX		
Talend	Open Studio	Offre Open-Source gratuite	Groupe Accor, GMF, Direction Générale de la Comptabilité Publique...
Pentaho	Data Integration	Offre Open-Source Dès 2000 €	T Mobile, Ericsson, General Electric...
Informatica	Power Center	Dès 50 000 €	Verizon Business, Société Générale, Sanofi Aventis...

Autres offres du marché

Editeur	Solution	Commentaire
Microsoft	SQL Server 2008	
Oracle	Oracle Warehouse Builder	
Cognos	Data Manager	
Clover.ETL	Clover.ETL (Open Source)	En gérant les processus légers (<i>threads</i>) de façon indépendante, Clover.ETL partage la charge de transformation sur différents processeurs et s'adapte ainsi aux multiples configurations matérielles
SAS	Data Integration Server	L'éditeur propose sa solution en mode intégré ou bien en SaaS et dispose notamment de connecteurs natifs pour les ERP SAP, PeopleSoft, Oracle et Siebel
SAP Business Objects	Data Integrator	Fonctionnalité de centralisation des métadonnées et un module complet de qualité des données (normalisation, correction...).
iWay Software	iWay DataMigrator	DataMigrator s'interface avec les principaux SGBDR du marché en utilisant une palette de connecteurs

Manapps, un groupe de conseil et de services basant son succès rapide sur une forte spécialisation dans les nouvelles technologies, le CRM, le MDM et la business intelligence a publié une mise à jour d'un benchmark très contesté sur les performances de cinq ETL. Fin 2008, une première volée de tests avait clairement placé l'ETL Open Source édité par Talend en tête devant les ETL d'IBM, celui de Pentaho et surtout celui d'Informatica. Le leader mondial de l'ETL se voyait assassiné dans les tests, ce qui allait déclencher un torrent de réactions sur le Web.

La nouvelle étude évalue les performances d'IBM Datastage Server 7.5, de Datastage PX 7.5, de Talend Open Studio 2.4.1, d'Informatica 8.1.1 et de Pentaho Data Integrator 3.0.0. Le benchmark porte sur onze tests différents, de complexité croissante avec des volumétries de données identiques au premier test et avec un simple PC de bureau sous xp comme machine de test. Cette fois, c'est Informatica qui s'en tire le mieux. L'américain obtient la meilleure note globale, devant Talend, les deux ETL IBM (Datastage PX devance Datastage server) et enfin Pentaho qui clôt le classement. Le benchmark est disponible gratuitement sur le site du cabinet de consulting.

Les perspectives et les challenges de l'ETL

L'implémentation de processus d'ETL efficaces et fiables comprend de nombreux challenges.

- Les volumes de données sont en croissance exponentielle et les processus d'ETL doivent traiter des quantités importantes de données. Certains systèmes décisionnels sont mis à jour de façon incrémentale, alors que d'autres sont rechargés dans leur totalité à chaque itération.
- Alors que les systèmes d'information se complexifient, la variété des sources de données s'accroît également. Les processus d'ETL doivent disposer d'une large palette de connecteurs à des progiciels (ERP, CRM...), bases de données, mainframes, fichiers, Services Web...
- Toutes les structures et applications décisionnelles (datawarehouse, datamart, applications OLAP...) présentent des besoins différents en termes de transformation de données, ainsi que des latences différentes.
- Les transformations des processus d'ETL peuvent être très complexes, et certaines, spécifiques au décisionnel, sont aussi requises.
- Le décisionnel se rapproche du temps réel, les datawarehouses et datamarts doivent donc être rafraîchis plus souvent sachant qu'il faut éviter les fenêtres de chargement.

Concernant l'avenir de l'ETL, il se situe peut être dans le cloud. En effet, Informatica s'est positionné sur cette niche de marché en proposant Informatica on-Demand. Nicholas Goodman, consultant et fondateur de Bayon Technologies s'est livré à un benchmark de la plate-forme Pentaho Data Integration déployée sur Amazon EC2. L'argument de ce choix de la plate-forme Amazon est financier et il est de taille : 4\$ facturés pour l'utilisation de quarante serveurs de petite taille (1 cœur + 1,7Go de RAM) et 2,80\$ pour les 20To d'espace de stockage. C'est dans ce mode de facturation d'Amazon EC2 que se cache l'intérêt du cloud pour l'ETL. Il n'est pas plus coûteux de louer cent machines pour un temps t qu'une seule machine durant cent fois ce temps. Pour peu que la montée en charge de l'ETL soit proche du linéaire, on va pouvoir mobiliser pour un traitement ETL un nombre de machines qu'il serait totalement inenvisageable d'avoir dans son propre datacenter. On assiste à un véritable changement de modèle qui pourrait très vite se développer dans un avenir proche.

Les nouveaux modes de stockage

Introduction

Dans cet axe nous allons nous intéresser principalement à la technologie « in-memory ». Nous verrons les différents acteurs de la BI qui utilisent cette technologie. Nous ferons ensuite un descriptif de cette technologie, un topo de ses avantages et enfin les solutions utilisées.

Les acteurs

Qliktech

QlikTech ouvre la voie à une nouvelle génération de solutions d'analyse décisionnelle, simples à utiliser, rapides et flexibles, qui permettent aux utilisateurs de manipuler instantanément les données grâce à une démarche intuitive, facilement compréhensible et très visuelle. QlikView 9, le produit phare de QlikTech, est construit sur une technologie brevetée d'association « in-memory » de nouvelle génération et de nombreuses fois primée. Il offre des solutions de reporting et d'analyse de l'information puissantes et visuelles, qui améliorent la prise de décisions quotidienne des opérationnels. Les solutions QlikView peuvent être déployées en quelques jours et les utilisateurs formés en quelques minutes, leur donnant ainsi la liberté d'être plus pertinents et créatifs dans leurs analyses.



SAP

SAP propose une gamme d'applications d'entreprises et de solutions de business intelligence afin de répondre aux besoins opérationnels en termes de gestion d'entreprise. Les logiciels de gestion SAP mettent en évidence les gisements d'amélioration des activités et permettent de les transformer en avantage compétitif. Au final, les solutions de gestion SAP apportent :



- la capacité à anticiper pour identifier de nouvelles opportunités ;
- l'agilité pour s'inscrire dans une dynamique d'amélioration continue ;
- les fonctionnalités nécessaires pour optimiser les opérations au quotidien ;
- des moyens permettant d'étendre les meilleures pratiques de gestion au delà de l'entreprise, à l'échelle de toute chaîne de valeur.

La technologie « In-memory »

Cette technologie repose sur le fait de stocker les informations, non plus sur le disque dur, mais dans la mémoire vive. La mémoire vive (RAM) ne dispose pas des limitations mécaniques et du temps de latence des disques durs. On obtient ainsi un gain de vitesse, de bruit, de consommation d'énergie et de fiabilité.

Dans un futur proche la technologie « **in-memory** » devrait remplacer les appareils de stockage mobile (disque dur externe, ...), et peut-être dans un futur plus éloigné, remplacer les disques durs internes. Actuellement, ce type de mémoire n'est pas encore concurrent des disques durs internes ou des serveurs utilisant une architecture RAID.

Les solutions

Qlikview

La technologie unique et brevetée d'association en mémoire de QlikView est le principe fondateur d'une plate-forme de business intelligence de nouvelle génération. Cette technologie permet de littéralement supprimer les problèmes et la complexité inhérents aux outils de BI traditionnels : l'incapacité d'élaborer des analyses rapidement, l'accès lent à des données et sources de données limitées et la nécessité de faire appel au service informatique de l'entreprise pour élaborer et modifier les analyses.

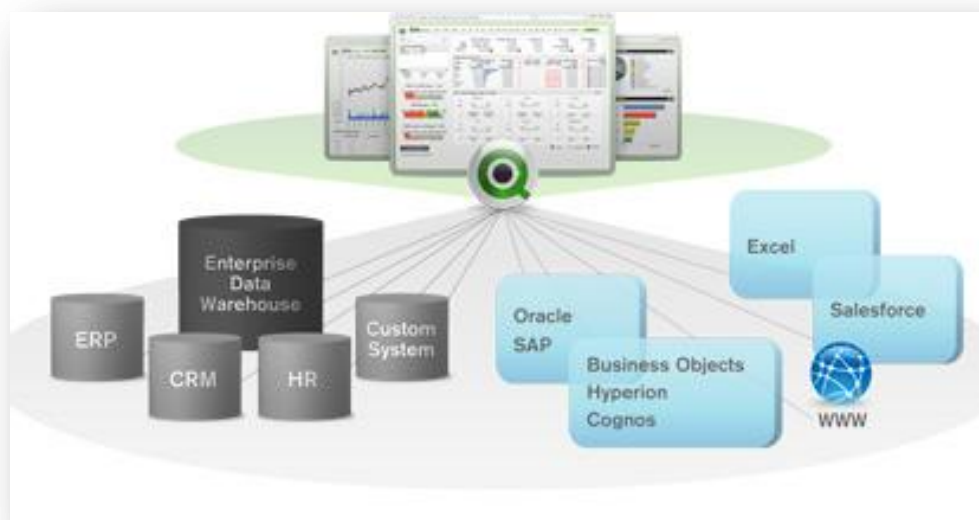


Schéma des fonctionnalités de QlikView

Puisque QlikView fonctionne entièrement en mémoire, il peut combiner des données à partir d'un nombre illimité de sources avec un niveau de performance élevé, et ce quel que soit le niveau de performance de chaque source. On profite ainsi d'un moyen plus rapide pour visualiser les relations entre les différents éléments de l'entreprise. QlikView le permet en associant des données en mémoire à partir de n'importe quelle source, par exemple : des systèmes transactionnels tels que SAP, Oracle, salesforce.com ; des entrepôts de données traditionnels tels que BusinessObjects, Cognos, Hyperion ; des bases de données telles que SQL Server, Oracle, MySQL qui optimisent vos applications personnalisées ; et des données locales issues d'Excel ou de fichiers plats. De plus, le fait que les données soient stockées en mémoire permet de contourner les limites des cubes sur disque traditionnel, ce qui rend toutes vos données accessibles rapidement.

Pour conclure, les avantages de Qlikview résident dans sa rapidité à obtenir des données ainsi que dans sa simplicité d'utilisation. De plus, Qlikview possède une interface graphique moderne et dynamique, ce qui rend le logiciel agréable à utiliser.

La solution de SAP

Bien qu'encore en développement, SAP tente aujourd'hui d'adapter la technologie « in-memory » à son système traditionnel. SAP prévoit d'une part une simplification de ses applications mais espère aussi « *remettre en cause le pouvoir des grandes bases de données, à commencer par celle du grand rival Oracle* » (D'après un article du LeMagIT Cf. sources). Il reste néanmoins à SAP à maîtriser la technologie et la faire adopter à son environnement. Bien qu'on trouve peu d'informations à ce sujet, je pense qu'il serait judicieux de suivre les évolutions de SAP dans ce domaine.

Conclusion

Actuellement QlikTech est la seule solution à utiliser la technologie « in-memory » dans le monde de la BI. Prochainement, on devrait voir apparaître un nouvel utilisateur SAP : sera-t-il aussi performant que QlikView ? Nous verrons cela dans les prochaines années.

La business intelligence en mode SaaS

Introduction

Avant ces dernières années, le monde des logiciels informatiques était confiné au modèle économique qui consiste à vendre un logiciel comme un produit, à proposer l'acquisition d'une licence (« on-premises ») nécessitant l'installation du logiciel chez le client. Mais depuis le début des années 2000, un nouveau modèle économique a fait son apparition sur le marché, le SaaS : Software as a Service (ou « on-demand ») proposant aux clients d'acquérir un logiciel périodiquement en y accédant par le biais d'internet, ce qui offre de nombreux avantages.

Les cabinets d'analystes sont très optimistes concernant l'utilisation du mode SaaS par les entreprises. Par exemple, selon Gartner, le SaaS devrait représenter 25% des revenus logiciels d'ici 2011. Le mode SaaS a connu ces dernières années un essor considérable, c'est pourquoi le domaine de la business intelligence n'est pas épargné et la majorité des grands éditeurs de BI propose aujourd'hui une offre en SaaS.

Cette partie dédiée à la business intelligence en SaaS présentera le marché actuel existant autour de cet axe, puis seront mis en évidence les apports de ce mode ainsi qu'un point sur la question sensible de la sécurité. Enfin, ayant une vue d'ensemble, une conclusion sera tirée concernant le mode SaaS.

Marché et acteurs

Les éditeurs et leurs solutions

Depuis l'apparition du modèle SaaS, son développement constant, et l'apparition de startups spécialisées dans la BI en SaaS, la plupart des grands éditeurs ont ajouté à leurs offres un outil en mode SaaS. La tendance est donc pour l'instant à l'ajout d'un outil de BI en SaaS à une gamme de produits en complément des outils classiques.

Editeurs		Solutions SaaS
		Oracle Hyperion On Demand
		SAS On Demand
		crystalreports.com SAP BusinessObjects BI OnDemand
		Cognos now!

Il faut savoir que ce tableau ne recense que les solutions SaaS des plus grands éditeurs de la business intelligence mais qu'il existe aussi bon nombre de « pure-players » ou sociétés spécialisées dans ce domaine.

Depuis 2008, le nombre d'offres de business intelligence en SaaS a considérablement augmenté et ce modèle, ainsi que son fonctionnement économique, sont de plus en plus entrés dans les mœurs. Cependant, aucun des éditeurs majeurs ne s'est vraiment imposé sur le marché, exception faite de SAP qui a pris une longueur d'avance avec ses offres « Business Objects On Demand ».

Les grands éditeurs communiquent peu sur le prix de l'abonnement à leurs offres SaaS. Cependant, parmi les acteurs plus secondaires du marché de la BI proposant des offres SaaS, le prix de l'abonnement est compris entre 30 et 100 € par mois et par utilisateur, en fonction des différentes fonctionnalités des applications choisies.

Les cibles

Les cibles de la business intelligence en mode SaaS sont clairement les entreprises de taille réduite ayant peu de moyens à attribuer au décisionnel. Ce sont donc les PME⁴ qui se tournent de plus en plus vers des solutions « on-demand » qui sont beaucoup moins onéreuses et qui leur offrent tout de même des services primordiaux, accessibles où ils le souhaitent grâce à internet. Le SaaS permet aussi aux entreprises d'assimiler leur utilisation d'une solution de business intelligence à un coût de fonctionnement et non plus à un investissement, il n'est plus obligatoire de déverser de fortes sommes pour la mise en place d'une solution classique.

⁴ Petites et Moyennes Entreprises

Les changements apportés par le SaaS

L'attrait que suscite le SaaS dans le monde du décisionnel s'explique par les nombreux avantages que celui-ci peut apporter aux clients et aux éditeurs/intégrateurs, bien que ces avantages soient plus prononcés pour les clients. De plus, avec le passage au « on-demand », les éditeurs se rémunèrent différemment.

Pour les éditeurs/intégrateurs

Pour les éditeurs et les intégrateurs de solutions de business intelligence, le SaaS permet surtout de cibler un nouveau marché, celui des PME qui n'ont pas un budget suffisant pour accéder à la BI « classique ». Cela peut donc leur permettre de toucher de nouveaux clients et d'élargir leur palette de revenus. De plus, par rapport aux solutions classiques, la mise en place des solutions en mode SaaS est considérablement simplifiée et donc très rapide.

Les sociétés proposant des solutions de BI en SaaS peuvent se rémunérer de deux façons différentes. La première consiste à proposer un forfait fixe quant à l'utilisation illimitée de la solution et la deuxième est d'être rémunérée en fonction de l'utilisation qui est faite de la solution (nombre d'utilisateurs, services activés, volumes consommés...).

Aujourd'hui, le pourcentage de revenu généré par la BI en mode SaaS est insuffisant et insignifiant en comparaison des revenus générés par les plateformes de BI classiques. La BI en SaaS représente moins de 5% du marché global de la business intelligence.

Pour les clients

Grâce au SaaS, il n'y a plus de contraintes matérielles ou logicielles pour le client, un simple poste avec un navigateur web suffit à accéder aux outils décisionnels. Les utilisateurs peuvent donc consulter et interroger les données de leur entreprise sans recourir à des logiciels onéreux et contraignants. De plus, mettre en place une solution de BI devient beaucoup plus rapide qu'auparavant puisque qu'il y a peu de changements à effectuer chez le client. Les coûts de maintenance s'en trouvent aussi très diminués, les besoins de déplacement chez le client étant réduits.

Ce sont donc les prix abordables, la facilité de déploiement et de maintenance qui sont au centre de l'intérêt des PME pour le mode SaaS.

Selon des études menées par différents cabinets d'analystes et notamment « Hurwitz & Associates », le coût total de possession (ou TCO : Total Cost of Ownership en anglais) d'une solution de BI en SaaS serait d'environ 70% inférieur à celui d'une solution classique.

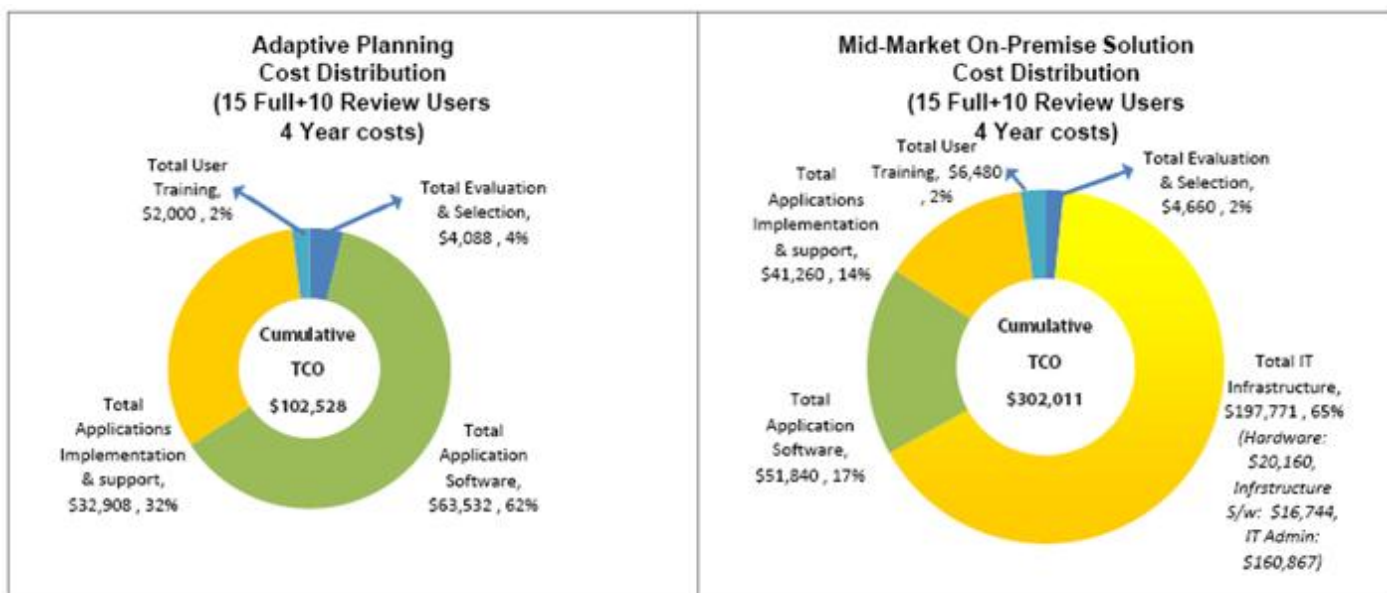
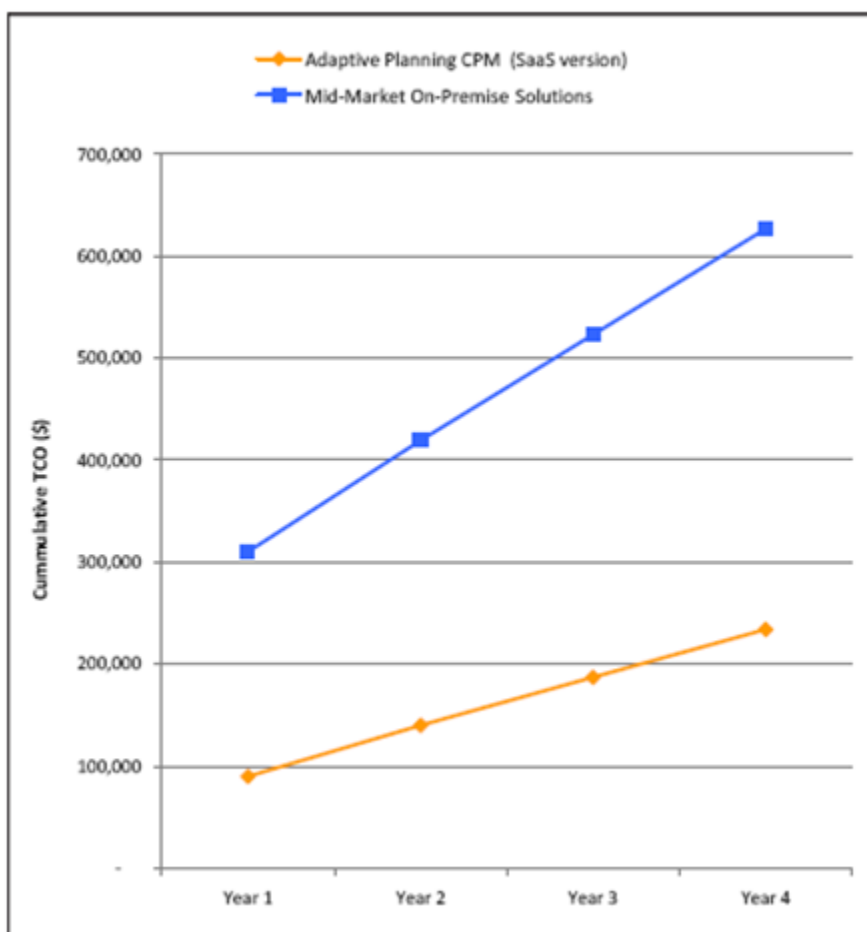


Schéma comparatif des coûts d'acquisition de solutions « on-demand » et « on-premises »

Source : Rapport du cabinet Hurwitz & Associates-“The TCO Advantages of SaaS-Based Budgeting, Forecasting & Reporting”



Graphique comparatif de l'évolution des coûts d'acquisition de solutions « on-demand » et « on-premises » sur les quatre premières années

Source : Rapport du cabinet Hurwitz & Associates-“The TCO Advantages of SaaS-Based Budgeting, Forecasting & Reporting”

Cette étude a permis de confirmer plusieurs points positifs concernant le choix d'une solution SaaS. Avec ce modèle, les coûts de souscription à l'outil se montent à 62% du coût total de la solution. Avec le modèle classique « on-premises » les coûts d'acquisition de la licence ne comptent que pour 17% du coût total de la solution. Ces coûts sont équivalents pour les deux modèles, or de nombreuses dépenses s'ajoutent à ceux-ci pour ce qui est du modèle classique, à savoir des coûts de conception, de maintenance, d'achat de matériel... ce qui rend ce modèle plus onéreux. Le SaaS permet aussi un déploiement plus rapide et plus flexible tout en ayant une fiabilité et des performances au niveau des solutions classiques.

Cependant, le Gartner invite les sociétés à être prudentes en ce qui concerne le choix de leur modèle de solution décisionnel car les avantages annoncés par les éditeurs de solutions SaaS peuvent se révéler moins importants que ceux promis dans l'offre commerciale.

Stockage de données et sécurité

Un des points les plus sensibles concernant le mode SaaS est la question de la sécurité du transfert et du stockage des données à l'extérieur de la société. Il est donc légitime pour un client de se poser des questions en ce qui concerne le choix d'une solution de BI en SaaS. En effet, on peut penser que les données seront beaucoup moins protégées si elles sont transférées via Internet plutôt que stockées directement au sein de l'entreprise.

Il faut savoir qu'aujourd'hui, pour ce qui est des solutions actuellement sur le marché, la plupart sont mises en place en partenariat avec un hébergeur, les données sont hébergées et maintenues par ce partenaire. Ce sont donc des sociétés spécialisées dans ce domaine qui gèrent ces flux et stockages de données et ils ont donc une expérience certaine en la matière.

De plus, si l'on se penche sur la réalité des architectures actuelles des systèmes d'informations, on s'aperçoit que l'ouverture et le partage des systèmes avec des tiers partenaires est une pratique avérée. Et, même si le risque zéro n'existe pas, avec leurs années d'expérience dans ce domaine, il est incontestable que la sécurisation des données n'a jamais été aussi fiable qu'aujourd'hui. Et pour en finir avec ces inquiétudes, les études démontrent qu'une application (transactionnelle ou décisionnelle) est, dans plus de 95% des cas, mieux sécurisée chez un hébergeur contractuellement responsable qu'au sein d'un système d'informations interne.

En ajoutant à cela l'avantage de la rapidité d'accès aux données, la question de l'externalisation des données ne devrait plus poser de problème.

Conclusion

Le marché de la BI en SaaS étant en phase d'expansion, il y a de nombreux acteurs proposant ce type de solutions que ce soit les grands éditeurs comme SAP ou des entreprises spécialisées dans le SaaS. Cela peut s'expliquer par le fait que ce modèle touche de nouveaux clients dans le monde du décisionnel, les PME. Nous avons pu voir qu'il y a de nombreux avantages pouvant inciter les PME à adopter une solution en SaaS avec un argument de poids : un prix raisonnable tout en bénéficiant des avantages d'outils classiques. De plus, le principal obstacle aux solutions SaaS étant la question de la sécurité, les partenaires de stockage de données sur internet disposent aujourd'hui d'assez d'expérience pour permettre aux éditeurs et intégrateurs de bâtir des argumentaires solides pour faire tomber cette dernière barrière.

Nous pensons donc qu'une solution SaaS peut être une bonne option pour compléter une gamme d'offres de BI classique dans le but d'étendre sa palette de clients. En revanche, miser sur une totale orientation vers le SaaS serait un pari très risqué aux vues des revenus générés actuellement par celui-ci. Bien que la BI « on-demand » connaisse une forte croissance au même titre que les autres domaines touchés par le SaaS, les revenus engendrés restent trop insuffisants comparés à ceux générés par une solution classique pour que le SaaS soit rentable.

La business intelligence mobile

Introduction

Le marché de la business intelligence ne cesse de s'étendre. En effet, en 2008, malgré des conditions économiques difficiles, la croissance des ventes de logiciels de business intelligence était de 20%, dépassant largement la plupart des autres marchés logiciels. Bien que ce marché s'étende, seulement 30% de potentiels utilisateurs de ces logiciels utilise la business intelligence en entreprise. Cela suggère que la BI au sein des entreprises est en train de changer et de se répandre. L'un des moteurs de cette évolution est le développement de solutions mobiles de business intelligence qui font de grands progrès en termes d'utilisabilité.

Cette partie du document vise à étudier les différents aspects de la business intelligence mobile. Nous allons tout d'abord lister les différents acteurs de ce domaine, puis nous présenterons des études des parts de marché des smartphones. Ensuite, nous nous poserons la question des contraintes de sécurité nécessaires pour protéger les données sensibles accessibles depuis les téléphones mobiles. Et enfin, nous essaierons de trouver une réponse à la question : ces solutions mobiles sont-elles réellement un plus pour les entreprises ou relèvent-elles plutôt du gadget ?

Marché et acteurs

On peut observer de nombreux acteurs présents dans le marché des logiciels de business intelligence pour téléphone mobile. Des éditeurs déjà existants adaptent leur solution pour smartphones, mais de nouveaux éditeurs spécialisés dans ces applications apparaissent également.

Voici un tableau récapitulatif des différents éditeurs présents sur le marché :

Société	Solution proposée	Compatibilité
BeeEye	BeeEye	<ul style="list-style-type: none"> Android (bientôt) BlackBerry® (bientôt) iPhone
 IBM	Cognos 8Go ! Mobile	<ul style="list-style-type: none"> BlackBerry® iPhone, iPad Microsoft Windows Mobile® Symbian S60 3rd Edition wireless
Information Builders	WebFOCUS	<ul style="list-style-type: none"> iPhone
Novedia	Novedia Decision	<ul style="list-style-type: none"> Java-enabled devices
Pentaho	Pentaho	<ul style="list-style-type: none"> iPhone
Prelytis	Prelytis	<ul style="list-style-type: none"> Tout appareil mobile ayant une connexion internet
 QlikTech	QlikView	<ul style="list-style-type: none"> Android iPhone, iTouch, iPad Java-enabled devices (BlackBerry®, Nokia)
 Regard	R-Survey™	<ul style="list-style-type: none"> Android BlackBerry® iPhone Microsoft Windows Mobile®
Roambi	Roambi	<ul style="list-style-type: none"> iPhone, iPad
 SAP Business Objects	SAP Business Objects Mobile	<ul style="list-style-type: none"> BlackBerry® iPhone, iTouch Microsoft Windows Mobile® Symbian
 Webalo	Webalo	<ul style="list-style-type: none"> Android (avril 2010) BlackBerry® iPhone (juin 2010) Java-enabled devices Microsoft Windows Mobile® Palm Symbian
 Editeurs spécialisés dans le développement d'applications de business intelligence pour smartphone.		

On observe donc qu'un grand nombre d'éditeurs développe aujourd'hui des solutions de business intelligence pour téléphone mobile. C'est un marché qui se propage car la mobilité est un facteur de plus en plus important pour les entreprises.

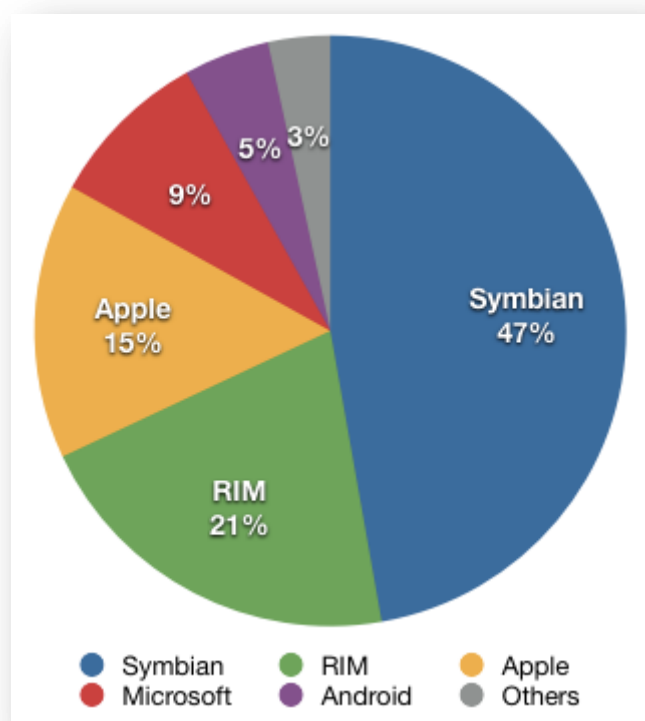
Ce type de solution vise des personnes souvent en déplacement et ayant besoin de prendre des décisions importantes dans de brefs délais. En effet, elles permettent d'avoir accès, depuis un téléphone mobile, aux informations présentes sur les serveurs de l'entreprise ; avec une interface toute aussi lisible et maniable que celle d'une solution de business intelligence classique. Les téléphones portables deviennent de plus en plus proches des ordinateurs, beaucoup sont tactiles, la plupart permettent aujourd'hui d'accéder à internet et de s'y déplacer avec aisance. Utiliser une application de business intelligence nécessite une grande interactivité et les éditeurs ont réussi à relever le challenge de rendre les solutions de business intelligence ergonomiques sur téléphone mobile. Le marché visé par les éditeurs est donc constitué de tous les décideurs qui sont très mobiles et doivent avoir la possibilité de consulter, voire même de modifier les données de l'entreprise à tout instant.

Marché des smartphones

Les smartphones sont des téléphones portables disposant de nombreuses fonctionnalités telles que l'agenda/calendrier, la navigation web, la consultation de courrier électronique, le GPS, etc... De part leurs nombreuses fonctionnalités, les smartphones sont très attractifs et ils prennent une part de plus en plus importante dans le marché des téléphones mobiles. Selon une étude faite par le cabinet Gartner, les ventes de smartphones devraient même dépasser celles de PC ; en effet, en 2013, il devrait y avoir 1,82 milliards de smartphones contre 1,78 milliards de PC.

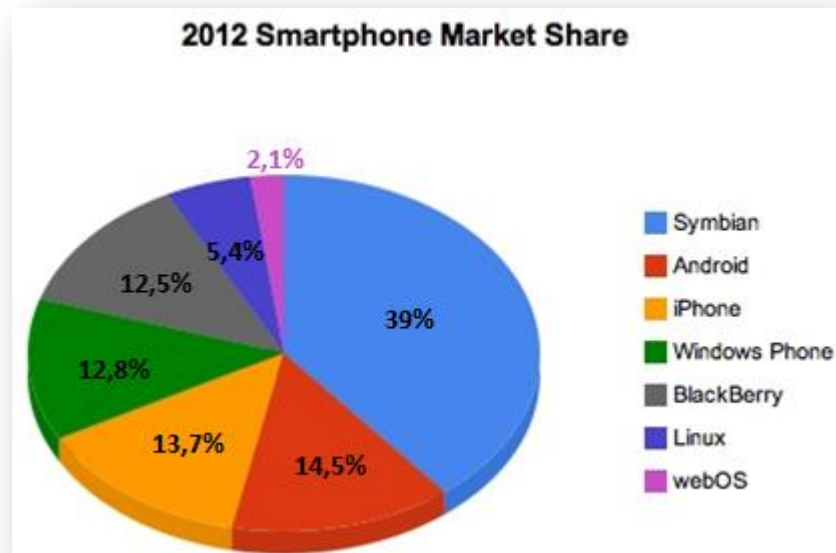
Il paraît intéressant d'étudier les parts de marché des smartphones dans le monde, par système d'exploitation. En effet, savoir quels systèmes d'exploitation prédomineront dans quelques années peut permettre de prévoir pour quels systèmes d'exploitation les éditeurs vont principalement développer leur application.

Pour l'année 2009, la répartition des parts de marché des différents systèmes d'exploitation a été la suivante :



Parts de marché des systèmes d'exploitation pour smartphones pour l'année 2009

Le cabinet Gartner a récemment réalisé une étude sur les parts de marché des différents systèmes d'exploitation⁵ de smartphones et a ainsi essayé d'établir ce que serait le marché en 2012. D'après cette étude, la part de marché pour les mobiles Android augmenterait de 12% d'ici à 2012, ce qui le positionnerait en deuxième position dans le Top 7 mondial des systèmes d'exploitation. La part de marché de Symbian devrait quant à elle chuter à 39%.



Prévisions des parts de marché des systèmes d'exploitation pour smartphones pour l'année 2012

Si l'on se fit aux chiffres précédents, les éditeurs de solutions pour téléphones mobiles auront donc intérêt à développer ou adapter leur solution pour des systèmes d'exploitation Symbian, Android et iPhone. Les BlackBerry® seront toujours présents sur le marché mais en net recul. La part de marché du système Microsoft Windows Mobile augmentera nettement mais restera en retrait par rapport à la prédominance de Symbian Android et iPhone.

⁵ Le système d'exploitation du BlackBerry est RIM

Les données transférées lors de l'utilisation de logiciels de business intelligence sont sensibles et doivent être protégées.

L'utilisation d'un VPN⁶ permet de faire une interconnexion avec le réseau local de l'entreprise. Cela repose sur un protocole permettant aux données de passer d'une extrémité à l'autre du réseau VPN en étant sécurisées car cryptées.

Malgré tout, plusieurs mesures élémentaires s'imposent pour préserver la sécurité des données :

- L'authentification : les solutions doivent être compatibles avec tous les modèles de sécurité et la procédure d'authentification doit être conforme au modèle en vigueur dans l'entreprise.
- L'autorisation : des droits d'accès doivent être attribués et adaptés aux différents utilisateurs. L'accès VPN des smartphones doit être restreint aux serveurs réellement nécessaires aux utilisateurs de téléphones portables.
- La sauvegarde des données et des communications sensibles : les données doivent être cryptées et les communications répondre à des critères très stricts tels que seuls les utilisateurs dûment authentifiés et autorisés peuvent décrypter les données. Les connexions venant des téléphones doivent être régulées par des systèmes de prévention des intrusions (IPS). Un serveur avec pare-feu où sont stockées les données de sécurité et autres informations propres à l'utilisateur doit être utilisé, séparé du reste de l'intranet.
- L'utilisation de mot de passe : le mot de passe doit être actualisé régulièrement et comprendre majuscules, minuscules et caractères alphanumériques.
- La neutralisation des appareils perdus ou volés : en cas de perte ou de vol, l'administrateur doit pouvoir empêcher l'accès aux données stockées en les supprimant puis en désactivant le mobile. Inconvénient : une fois l'appareil désactivé, il est impossible d'effacer les données à distance.

C'est en additionnant les fonctions de sécurité des terminaux mobiles et celles des solutions de business intelligence que les entreprises protégeront correctement leurs données. Avec de bonnes pratiques et une prévention auprès des utilisateurs, la sécurité des données consultées depuis un terminal mobile peut tout à fait être assurée sans représenter de contrainte majeure.

⁶ Virtual Private Network = réseau privé virtuel

Quelque soit le secteur d'activité, la mobilité est aujourd'hui une réalité que les entreprises ne doivent pas négliger. De nouveaux applicatifs accessibles depuis le web se développent alors, permettant aux entreprises d'accéder aux informations dont elles ont besoin, même en déplacement. Parallèlement à cela, les smartphones se développent également de plus en plus, permettant d'accéder au web, à ses e-mails, accroissant ainsi la facilité d'accès aux informations depuis son mobile, soit n'importe quand et depuis n'importe où. Aujourd'hui, la business intelligence ressent elle aussi le besoin de se développer pour les terminaux mobiles, permettant aux utilisateurs d'avoir accès depuis leur mobile, à des données calculées en temps réel. L'ergonomie des solutions actuellement développées est tout à fait propice à l'utilisation mobile (tableaux interactifs, alertes, etc...). Le marché de la mobilité d'entreprise devrait donc connaître une très forte progression au cours des prochaines années et, selon le cabinet d'analyse Gartner, « le nombre d'applications mobiles déployées par les sociétés auprès de leurs collaborateurs va continuer à croître de 30 % par an d'ici 2011 (probabilité de 0,7) ».

Malgré cet engouement pour les applications mobiles, on peut se demander si des applications de business intelligence pour téléphones mobiles seraient réellement un plus pour les entreprises. En effet, ces solutions permettent aux décideurs d'avoir accès tout le temps et partout aux données actualisées de sa société, mais est-ce un réel atout pour les entreprises ? Un décideur a-t'il réellement besoin d'accéder aux données de l'entreprise en dehors de celle-ci ? Il est difficile de répondre à ces questions sans interroger les personnes concernées. D'un point de vue extérieur, ces nouvelles solutions paraissent révolutionnaires mais toute nouvelle technologie ne donne t'elle pas cette impression à l'aube de sa sortie ? Peut-être alors qu'effectivement, ces nouvelles applications se révéleront être un peu gadget au fil du temps.

Conclusion

Pour conclure, nous avons vu qu'il existe de nombreux acteurs présents dans le domaine des applications mobiles de business intelligence. Ces acteurs peuvent être des éditeurs de solutions de business intelligence déjà existants, comme SAP ou encore Cognos, qui développent une solution mobile de leur application phare. Ces entreprises peuvent également être des éditeurs spécialisés, c'est-à-dire qu'ils proposent uniquement des solutions mobiles de business intelligence. Ces solutions mobiles sont faites pour être utilisées sur des smartphones, nous avons donc ensuite étudié le marché des systèmes d'exploitation de smartphone et essayé d'anticiper ce qu'il serait d'ici à 2012 afin de savoir vers quels systèmes d'exploitation les éditeurs de solution mobile devraient se tourner. Par la suite, nous nous sommes posés le problème des contraintes de sécurité imposées par les solutions mobiles et il est ressorti qu'elles paraissaient tout à fait surmontables. Enfin, à la question des apports de la mobilité, il est vrai que les avantages semblent nombreux étant donné le contexte de mobilité actuel, mais seuls des témoignages de professionnels et l'avenir pourront nous dire si ces solutions apportent un réel plus aux entreprises.

Les moteurs de recherche

Introduction

Les outils de recherche peuvent se classer en trois grandes catégories :

- les moteurs de recherche sur Internet, qui sont les plus communément utilisés ;
- les logiciels de recherche installés sur un ordinateur personnel, comme Exalead Desktop, Google Desktop et Copernic Desktop Search ;
- les outils dédiés à la recherche de données dans un réseau d'entreprise.

Ce sont ces derniers outils qui nous intéressent dans cette partie.

Qu'est-ce qu'un moteur de recherche ?

Un moteur de recherche est un outil permettant de retrouver des ressources associées à des mots quelconques. Chaque document identifié est alors indexé dans une base de données, accessible ensuite par les utilisateurs à partir de mots-clés.

Pourquoi un moteur de recherche dans une entreprise ?

L'explosion du nombre de contenus de formats divers (données, informations non structurées, images, vidéos...) disponibles dans les entreprises les pousse à s'équiper de moteurs de recherche en interne.

Selon une étude menée par MARKESS International, 49% des organisations ont déjà recours à un moteur de recherche d'entreprise, et 18% envisagent son utilisation d'ici à 2010. Ces moteurs de recherche sont en majeure partie intégrés aux postes de travail ou aux outils de gestion électronique des documents, mais ils sont, dans un nombre grandissant d'entreprises, capables de couvrir à la fois les contenus internes et externes de l'entreprise, ou encore intégrés aux outils de gestion de contenu ou aux solutions décisionnelles.

Les moteurs de recherche peuvent être abordés au travers d'autres problématiques, comme les médias de restitution ou la qualité des données présentées (MDM).

L'importance du moteur de recherche

Pas de business intelligence et de réactivité sans un bon moteur de recherche. Il faut désormais pouvoir tirer parti de toutes les sources de données de l'entreprise : PGI, CRM, GED, intranets, système de collaboration...

Puis les besoins de performance, de sécurité et d'extension de la couverture des formats ont augmenté. Le moteur de recherche devient le point clé d'accès aux données et informations de l'entreprise. Les opérationnels ont besoin d'un vrai service de recherche, accédant à un maximum de sources existantes où qu'elles se trouvent (poste de travail, web, GED, mails, PGI, applications métiers, gestion des connaissances, bases de données clients, factures, forums, wikis, réseaux sociaux internes...). Ce service doit s'intégrer aux modes et contrôles d'accès ainsi qu'aux systèmes d'archivage existants. Et cela pour un patrimoine informationnel qui double tous les dix-huit mois. Enfin, pour les utilisateurs finaux, il doit être aussi simple et rapide que Google l'est sur la toile.

Et d'abord qu'est-ce qu'un bon moteur ? Il doit idéalement combiner une analyse statistique et sémantique. La première apprécie l'importance des mots selon leur occurrence et leur emplacement dans les textes. L'autre reconnaît leur sens et sait donc associer par exemple les mots déplacement, avion et voiture. Tous les éditeurs sont bons sur l'analyse statistique, mais plus inégaux sur l'analyse sémantique.

Le marché et les acteurs

Gartner prévoit que le marché de l'Enterprise Search sera de 1,5 milliards de dollars en 2012, contre 860 millions en 2007, soit un taux de croissance de 11,4 % en moyenne par an.

Les clients des moteurs de recherche sont très variés. Quelque soit l'entreprise, elle pourra intégrer dans son réseau une solution de recherche.

Les caractéristiques des trois acteurs principaux ont été regroupées dans le tableau ci-dessous :



	Exalead	Endeca	Sinequa
Siège social	Paris	Boston	Paris
Fondé en	2000	1999	2000
Nb employés	160	500	50 en 2008 (*3 en 5 ans)
Produit(s)	Exalead CloudView Search	Endeca's Information Access Platform	Sinequa CS : solution d'accès à la connaissance pour l'entreprise
Formats supportés	320		> 300
Volumétrie	> un milliard de documents	> un milliard de documents	> un milliard de documents
Rapidité	800 req/sec sur 15 millions de docs		400 requêtes par seconde (les Pages Jaunes européennes)
CA	13 millions d'euros en 2007 15,5 en 2008	100 millions de dollars en 2007	7 millions d'euros en 2008
Nb clients	> 250 (site Exalead)	env. 500 (janvier 2008)	> 200 (mars 2009)
Références	Via Michelin, Air Liquide, 118 218, Société Générale, BNP Paribas	Boeing, Ford, JBM, Toshiba	Bouygues Construction, Arkema, la DGA, EADS

Selon les cabinets d'analyse Gartner, IDC et Forrester, Endeca est actuellement le leader du marché de la recherche en entreprise et de l'accès à l'information. D'autres acteurs s'illustrent grâce à leur outil de recherche.

Autonomy

Autonomy est une entreprise anglaise implantée partout dans le monde. Les technologies des solutions développées cherchent à donner du sens aux informations non structurées (texte, voix ou vidéo) pour ensuite pouvoir exécuter des traitements automatiques. Leur moteur de recherche se nomme Idol K2, il apparaît comme le plus riche fonctionnellement.

Google

Au travers de sa branche dédiée aux entreprises, Google propose sa solution de recherche en entreprise : Google Search Appliance.

Vivisimo

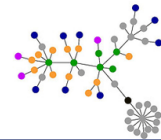
La solution Velocity Search Platform 7.0 de Vivisimo apporte pour sa part la recherche conceptuelle, fonction du profil de l'utilisateur et de son contexte, ainsi que la recherche sociale et la gestion avancée de la catégorisation automatique. Cet éditeur américain, cofondé par le français Jérôme Pesenti, a notamment équipé Airbus, Procter & Gamble, la National Library of Medicine, Cisco, USA.gov et l'US Air Force. Sa technologie peut être librement testée via le méta-moteur Clusty.com, qui peut interroger simultanément neuf moteurs différents et qui catégorise automatiquement les résultats par thématique.

Fast (Microsoft)

Fast Search & Transfer a lui aussi une couverture fonctionnelle très large. Son créateur, une jeune pousse norvégienne de sept cents cinquante personnes désormais, est également l'auteur du moteur de recherche web AllTheWeb, d'abord racheté Overture, puis par Yahoo!. Il équipe plus de deux milles six cents clients (AOL, Dell, IBM, Wanadoo, Vodafone, Lexis-Nexis, Reuters, T-online, US Army...). Microsoft l'a racheté en avril 2008 pour 1,2 milliard de dollars. Logique : ceux qui se heurtaient aux limites du moteur de SharePoint Server le remplaçaient généralement par un moteur plus puissant comme Fast. A terme, Microsoft compte évidemment en faire une plate-forme de recherche intégrée pour l'ensemble de son environnement.

Omnifind (IBM)

Omnifind a été taillé pour des millions de documents et des milliers d'utilisateurs. Moteur sémantique de plus de soixante langues, il analyse les contenus, les sentiments et les menaces dans les serveurs Sharepoint, Filenet, Documentum, MS Exchange, Lotus Notes et Websphere.



Qu'est-ce que le web sémantique ?

Le web sémantique se base sur la standardisation des métadonnées des ressources web. Il va permettre de décrire les relations entre personnes, à l'aide de langages destinés à structurer ces vocabulaires. Ce principe repose sur des standards déjà bien définis tels que RDF et OWL. Ils sont issus du W3C, organisme chargé de promouvoir la compatibilité des technologies du World Wide Web. Ces standards concrétisent la notion d'intelligence collective.

En quoi ces technologies vont-elles modifier les moteurs de recherche ?

Cela va sans doute prendre beaucoup de temps, mais les développeurs et créateurs de ressources vont de plus en plus utiliser les métadonnées standardisées du web sémantique. Les moteurs de recherche doivent alors mieux prendre en compte ces nouvelles informations pour pouvoir les exploiter.

Une étude du Gartner datée de juillet 2009 indique que les technologies du web sémantique ne seront exploitées pleinement que dans une dizaine d'années.

Certains éditeurs ont déjà bien compris l'intérêt de ces nouvelles technologies, à l'image d'Exalead qui investit dans ce domaine au travers de son pôle innovation. Son moteur de recherche web intègre la prise en compte des métadonnées.

L'éditeur Sinequa s'intéresse également à ces nouvelles technologies, sa solution Enterprise Search crée des liens sémantiques entre les données dans un contexte métier pour aider les employés à trouver l'information et à prendre des décisions plus rapidement.

« Faire réfléchir » les ordinateurs : c'est l'idée de Fabrice Lacroix, président d'Antidot, éditeur français de solutions d'accès à l'information, outils de recherche et de navigation. Ayant emboîté très tôt le pas de la sémantique, Antidot se pose à présent comme l'un des acteurs les plus avancés du domaine. Cette affirmation se vérifie au travers de leur important catalogue client.

Le futur des moteurs de recherche

L'élément central sur lequel était construit le système d'information de l'entreprise 1.0 était la base de données relationnelle mais cela implique souvent un système lourd et coûteux. De plus, les outils de business intelligence manquent d'agilité ; et enfin, les données non structurées sont devenues de plus en plus métiers et critiques.

Parallèlement à cela, l'info-explosion des données non structurées (les mails et les fichiers qui sont stockés « à plat », sans modèle relationnel) a fait progresser les technologies de recherche.

Le CEO de Sinequa indique que dans 90% des besoins, le moteur de recherche fait plus souple et mieux que l'outil de BI.

En plus de l'intégration des technologies du web sémantique, les moteurs de recherche devraient prendre en compte d'autres problématiques.

Dans un futur proche, la problématique de la gestion des données de référence (MDM) deviendra inévitable. C'est alors qu'entre en jeu l'Enterprise Search Bus, un bus de moteur de recherche intelligent capable de mettre en cohérence l'ensemble des données en les rendant « recherchables » et navigables par des axes métier.

Les futurs moteurs de recherche devront également s'appuyer sur des solutions de Master Data Management afin de présenter des données d'une qualité et d'une fraîcheur toujours meilleures.

Certains éditeurs proposent par ailleurs leurs services en mode SaaS, à l'image des produits Exalead On Demand. Il s'agit là aussi d'une évolution possible des solutions de recherche à court et moyen terme.

Les médias de restitution

Introduction

L'indexation est à la base une pratique documentaire dédiée aux textes, et dont l'objectif est de générer pour un document, une description sous forme d'index permettant de le retrouver dans un ensemble. Le choix des caractéristiques du document et du langage de description détermine la pertinence des index. Ces questions sont au centre des recherches actuelles sur l'indexation automatique de documents audiovisuels. Désormais, on arrive à indexer d'autres médias comme l'image, le son et la vidéo. Ensuite, grâce à des moteurs de recherches, en entrant des mots clés ou des couleurs, il est possible de retrouver les médias indexés.

L'information non indexée (non structurée) est de plus en plus présente sur le net. Aujourd'hui, 80% des informations d'une entreprise ne sont pas structurées et selon Gartner, le volume de ces données double tous les mois. Ces informations étant chères à traiter, le défi pour l'entreprise moderne est d'essayer de comprendre et d'extraire la valeur qui se trouve dans cette vaste mer de données.

Dans une première partie, nous allons voir les différents acteurs par média et ensuite les solutions qu'ils proposent.

Pourquoi l'indexation ?

Mais pourquoi est-il autant important de faire de l'indexation ? Voici ci-après 3 apports de l'indexation.

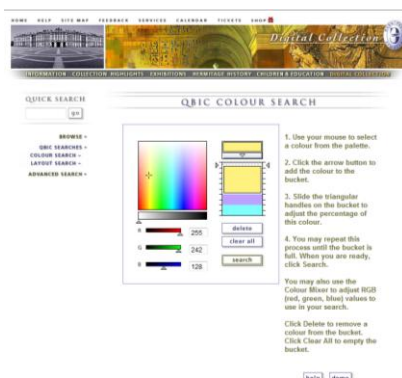
- Permettre la découverte de connaissances en organisant et en classifiant de façon dynamique l'actif informationnel en dossiers consultables pertinents dans le contexte, afin de permettre à l'utilisateur de prendre des décisions opportunes et de résoudre les problèmes rencontrés dans le monde réel.
- Trouver des informations pertinentes sur le plan conceptuel, et les convertir en connaissances quelle que soit la langue dans laquelle elles sont conservées et les perspectives de recherche de l'utilisateur final.
- Récupérer des résultats de recherche exhaustifs et pertinents, y compris lorsque les termes de la recherche sont différents de la source, tout en y étant apparentés.

Les acteurs

Acteur	Solution	Média traité			
		Image	Son	Vidéo	Carte
Google	Gaudi	-	X	X	-
INRA	IMEDIA	X	-	-	-
IBM	CQBI	X	-	-	-
Business Geografic	Aigle	-	-	-	X
Exalead	CloudView	X	X	X	-
Groupe Quaero	Quaero	X	X	X	-
IT channel.info	LTU	X	-	-	-
Seigneurie	Chromatic	X			-
Université de Cambridge	Autonomy	X	X	X	-
Google	Google Location Services	-	-	-	X
NewPhenix	NewPhenix	X	-	-	-

Les solutions

IMEDIA de l'INRIA



Le musée de l'hermitage utilise IMEDIA comme moteur de recherche

L'objectif premier de l'équipe-projet IMEDIA est de développer des méthodes d'indexation par le contenu, de recherche interactive, et de navigation dans des bases d'images, dans un contexte multimédia. Pour ce faire, IMEDIA traite aussi bien des bases d'images « génériques » (ex : web) que des bases d'images « spécifiques » à un domaine d'application ciblé (visages, images médicales...). Ces deux catégories relèvent respectivement de la recherche d'images et de la reconnaissance d'objets. En réalité, la recherche d'images est une problématique plus vaste qui englobe la reconnaissance d'objets et intègre les interactions avec l'utilisateur. Plus généralement, l'équipe-projet

IMEDIA déploie ses efforts de recherche, de collaboration et de transfert pour répondre au problème complexe de l'accès intelligent aux données multimédias dans sa globalité.

Actuellement, on ne trouve aucune information récente sur la toile. Ce projet a été soit abandonné soit a changé de nom. Par ailleurs, l'équipe de l'INRIA travaille sur le projet Quaero (indexation multi-médias) avec de nombreux collaborateurs (Cf. Quaero, un peu plus loin).

Google Audio Indexing (Gaudi)

Google Audio Indexing est en fait une nouvelle technologie de Google qui permet aux utilisateurs d'effectuer de meilleures recherches vidéo et regarder des vidéos YouTube à partir de divers canaux.



Exemple d'utilisation de Gaudi sur le site Youtube avec le thème des élections présidentielles américaines

C'est une technologie qui permet de trouver, dans une vidéo, un mot que la personne a dit. Elle permet à l'utilisateur d'accéder directement à la partie de la vidéo où ces mots sont prononcés. Google réalise ici un grand pas dans le monde des moteurs de recherche.

Pour le moment, les tests sont effectués sur la prochaine élection présidentielle où est basé Google ; mais ils ont déjà prévu d'améliorer le service pour le proposer ensuite à d'autres vidéos.

Il y a quelque mois, on pouvait tester cet outil sur les vidéos des élections présidentielles américaines, mais désormais tous les liens sont morts.

Aigle de Business Geographic



Aperçu de l'interface de Business Geographic

Business Geographic, filiale du groupe CIRIL, est l'inventeur du concept de « Géo Intelligence ». Editeur de la technologie Aigle™, Business Geographic présente GeoBI, la solution BI dédiée au reporting géographique en environnement Business Objects. GeoBI est la solution offrant un véritable lien dynamique entre votre portail Business Objects WEBI/DASHBOARD ou votre environnement client/serveur BO et vos données géographiques. Alliant puissance et simplicité, GeoBI s'impose comme la solution leader sur le marché du Géo-Décisionnel.

GeoBI est un véritable outil de reporting et d'analyse cartographique en environnement Intranet/Extranet et Internet. GeoBI apporte une nouvelle dimension à l'information en offrant la possibilité de consulter, d'analyser et de partager des tableaux de bords géographiques.

Dans un article de 'directions magazine' publié le 18 Mai 2010 (Cf. Annexe), Business Geographic a récemment ajouté un volet mobile développé pour iPod®. L'application assure la mise à jour de données en temps réel, afin que la réactivité et la crédibilité de l'équipe sur le terrain soit améliorées. «The iPod interface is very intuitive; it really makes the field operations more efficient », a déclaré Dany Bouchard, Président de DBx GEOMATICS.

Exalead CloudView

CloudView est une solution pertinente et reconnue pour le développement d'applications métier. La business intelligence nouvelle génération met l'utilisateur au centre des applications et délivre une vue à 360 degrés sur l'information de l'entreprise. Les Search Based Applications (SBA) comme celles reposant sur CloudView vont plus loin, en intégrant les données non structurées à forte valeur ajoutée : fichiers, courriels, Web...

Celles-ci garantissent un accès instantané aux données et s'intègrent de façon non intrusive au système d'information existant. Elles préfigurent ainsi les nouveaux systèmes d'informatique décisionnelle rich media qui vont rendre l'entreprise plus agile et réactive.

D'après un article de 'directions magazine' publié le 29 Avril 2010 (Cf. Annexe), Exalead est devenu le partenaire de Micropole-Univers. Fort de plus de vingt années d'expertise opérationnelle dans les domaines de la business intelligence, du Web & IT, des ERP et du CRM, Micropole-Univers est reconnu par ses clients pour sa connaissance de leurs métiers et sa maîtrise des technologies. C'est dans cette logique d'innovation constante que le groupe a intégré la plateforme Exalead CloudView à son portefeuille de solutions, notamment pour ses possibilités de développement d'applications métiers facilitant l'agilité et la réactivité des entreprises, à moindre coût.

Quaero

Quaero (« je cherche » en latin) est un programme de recherche et d'innovation destiné à développer des « outils intégrés de gestion des contenus multimédias », dont des extensions multimédias pour des moteurs de recherche de nouvelle génération qui permettront de rechercher par le contenu non seulement du texte, mais aussi des images, du son et de la vidéo.

Ce projet est développé par vingt-cinq partenaires : grands groupes, PME, laboratoires de recherche, établissements publics.



Les collaborateurs du projet Quaero

« D'importants efforts de recherche sont nécessaires pour que l'ordinateur effectue ce que les humains font naturellement : reconnaître les visages et les voix, comprendre ce qui se dit et la façon de le dire... Quaero a l'ambition d'élaborer des solutions nouvelles pour faire émerger de nouveaux services d'accès à l'information. Ainsi, dans le domaine du langage parlé, Quaero vise à identifier la langue utilisée et le locuteur, à transcrire la parole, ou à traduire automatiquement d'une langue à l'autre. » dit Joseph Mariani, directeur de l'Institut des technologies Multilingues et Multimédias de l'Information (IMMI).

Fin mars 2010, le site de l'Elysée a été lancé avec, comme moteur de recherche, Voxlead « Inspired by Quaero ». Une première en France dans l'indexation vidéo !

LTU

LTU technologies est à la pointe des technologies existantes dans le domaine de la détection vidéo (« video matching ») tant en termes de performance que de précision. Les acteurs du marché, et particulièrement les détenteurs de droits, les sites de partages de vidéos, les réseaux sociaux sont demandeurs de solutions pour gérer l'explosion des contenus multimédias, ainsi que l'utilisation frauduleuse toujours plus importante de contenus protégés.

LTU applique son expertise unique en analyse d'images pour rechercher des correspondances entre vidéos, quel que soit leur format. Dans un premier temps, des images sont extraites de vidéos de référence et leur ADN est calculé. Ensuite, lorsqu'une vidéo requête est soumise au système, des images en sont extraites. Ainsi, les ADN des images de la requête peuvent être comparés aux ADN de la référence. Lorsqu'un nombre suffisant d'images en commun est trouvé entre la requête et une vidéo de la référence, une alerte est levée. Des algorithmes de matching très pointus permettent d'éviter les faux positifs tout en supportant de nombreuses transformations.

LTU technologies, pionnier des technologies de recherche et de reconnaissance d'images, a ouvert en avril 2010 une API permettant de créer des applications sur mesure mettant en œuvre des fonctionnalités de recherche par l'image. Afin de démontrer l'efficacité et la facilité d'intégration de son API de recherche visuelle, la société présente en parallèle lookthatup, une application iPhone de recherche par l'image. Véritable vitrine de la technologie de recherche et de reconnaissance d'images de LTU, cette application a été développée uniquement à partir de plateformes open source et de l'API LTU. Elle permet aux utilisateurs de reconnaître certains objets en les prenant en photo, puis de les bookmarker ou les partager sur les réseaux sociaux. Elle permet également d'ajouter facilement du contenu à la base de données d'images reconnaissables de lookthatup.

Chromatic

Chromatic est un outil pour les architectes et maîtres d'œuvre offrant une classification fine des couleurs proposées par le fabricant :

- Identification rapide par hauteur de ton désiré ;
- Répartition des couleurs par cycle chromatique, allant des tons pastel, aux tons moyens, puis aux tons forts et enfin aux tons gris.



Panel de couleur
proposé par Chromatic

Autonomy

Autonomy Corporation plc (LSE : AU. ou AU.L) est le leader mondial sur le marché du logiciel d'infrastructure d'entreprise et des technologies dites de « meaning-based computing ». La technologie Autonomy permet la compréhension conceptuelle et contextuelle de toutes les données électroniques y compris les informations non structurées, qu'il s'agisse de données texte, email, voix ou vidéo. Le logiciel Autonomy démontre son efficacité dans tous les domaines stratégiques d'une organisation : accès à l'information, IE, GRC, KM, solutions pour centres d'appels, gestion de contenus multimédias, solutions de conformité et traitement des litiges, applications de sécurité. Il est reconnu par les analystes comme le leader du marché professionnel de la recherche d'informations.

Le Groupe Autonomy se compose de :

- ZANTAZ, leader de l'archivage, de la détection d'informations au format électronique et de la gestion proactive des risques liés aux informations (Information Risk Management, soit IRM) ;
- Cardiff, l'un des plus grands éditeurs de solutions intelligentes de traitement des documents ;
- etalk, fournisseur réputé de produits de centres d'appels d'entreprises ;
- Virage, pionnier des technologies de gestion des contenus multimédias, de sécurité et de surveillance.

Autonomy a été fondée sur une vision de changer radicalement la manière dont nous interagissons avec l'information et les ordinateurs, s'assurer que les ordinateurs structurent notre monde, plutôt que l'inverse.

«L'essence du logiciel Autonomy réside dans sa capacité à extraire les concepts de base de données non structurées. La manière dont elle y parvient est sans doute une voie à suivre. » Whiting Peter, UBS.

NewPhenix

NewPhenix développe des logiciels d'analyse, d'indexation, de recherche et de classification de l'information multimédia et multilingue. Ils sont modulaires afin d'être intégrés et s'appuient sur des technologies linguistiques/sémantiques et pixéliques.

Google Location Services

Grâce au service Google Location Services, le navigateur peut transmettre vos coordonnées géographiques aux sites Web afin que vous puissiez trouver les informations les plus pertinentes et utile. Un site Web sera capable de vous demander de partager votre localisation de sorte qu'en cherchant simplement le mot « pizza » le site vous donnera les réponses dont vous avez besoin... aucune information ou saisie de texte supplémentaire n'est nécessaire. Un autre exemple : vous cherchez votre route pour vous rendre quelque part, le site Web saura où vous vous trouvez, ainsi vous n'aurez à indiquer que votre destination.

Ce service est totalement optionnel et le navigateur Web ne partage pas votre localisation sans votre permission ; ceci est réalisé avec le plus grand respect de votre vie privée.

Si vous êtes d'accord, le navigateur rassemblera les informations sur les points d'accès sans fil alentour et l'adresse IP de votre ordinateur. Alors, il enverra ces informations au fournisseur de service de géolocalisation par défaut, Google Location Services, pour faire une estimation de votre localisation. Cette estimation de localisation est alors partagée avec le site Web qui l'a demandée.

Si vous n'êtes pas d'accord, le navigateur ne fera rien.

Par défaut, le navigateur utilise Google Location Services pour déterminer votre localisation en envoyant :

- l'adresse IP de votre ordinateur ;
- des informations sur les points d'accès sans fil environnants ;
- un identifiant client aléatoire, assigné par Google, et qui expire après deux semaines, enregistré sous la forme de cookies.

Conclusion

Depuis que les contenus audiovisuels s'échangent électroniquement des données, les techniques d'indexation automatique se développent :

- pour le texte ;
- pour l'image ;
- pour le son ;
- pour la vidéo ;
- pour les cartes ;

et les formats d'index se standardisent.

Néanmoins, l'augmentation massive de documents ne cesse de croître et il est de plus en plus important de faire le tri entre l'utile et le futile.

Master Data Management

Introduction

La gestion de la multitude de données différentes existant dans une entreprise est coûteuse et complexe. La gestion des référentiels métiers (master data management en anglais, MDM) a pour but de référencer, contrôler, synchroniser toutes les données à travers l'entreprise (par exemple les données clients et produits), de façon à ce qu'elles soient précises, conformes, exploitables par les systèmes et leurs utilisateurs. Le master data management est un ensemble de disciplines, technologies et solutions utilisées pour assurer l'exactitude, la cohérence et l'unicité des données au sein de l'entreprise.

Pourquoi le MDM ?

Les conséquences de la faible qualité des données sont nombreuses et coûtent souvent cher aux entreprises. Les conséquences s'avèrent parfois fatales aux organisations dont les données sont de mauvaise qualité :

- **Insatisfaction des clients** : les clients victimes de données incorrectes pardonnent difficilement.
- **Non-conformité des chiffres publiés** : des erreurs dans les données comptables entraînent généralement de lourdes sanctions (amendes et peines). Et ce, même si cette situation ne résulte pas d'une manipulation volontaire.
- **Dévalorisation de l'image de l'entreprise** : en conséquence du point précédent, des erreurs portant sur des données diffusées à l'extérieur de l'entreprise détériorent rapidement l'image de celle-ci, et handicapent son expansion commerciale. De plus, l'impact boursier des sociétés cotées est immédiat, entraînant la chute rapide du cours de l'action.
- **Perturbation du fonctionnement opérationnel** : les données incorrectes au sein de l'entreprise ralentissent le fonctionnement opérationnel, et rendent les décisions difficiles ou incertaines. Cela génère aussi du stress et de la démotivation chez les collaborateurs. De plus, le surcroît de travail engendré par des erreurs de données n'est jamais le bienvenu. Enfin, la perte de confiance dans les informations manipulées n'est propice ni à l'initiative, ni à un climat sain.
- **Erreurs de stratégie** : définir une stratégie nécessite plus de temps si les informations utilisées sont de mauvaise qualité ou tout simplement erronées.
- **Augmentation des coûts** : les impacts énoncés précédemment engendrent évidemment des surcoûts. Les situations variant d'une entreprise à une autre, il est très difficile d'estimer le coût de la non-qualité. Certains experts estiment cependant que ce coût caché peut représenter plus de 10 % du chiffre d'affaires, ce qui ampute de façon conséquente la rentabilité.

La gestion de l'information est un facteur clef de la compétitivité de l'entreprise. Une information cohérente, intègre, fiable, accessible et sécurisée permet entre autres :

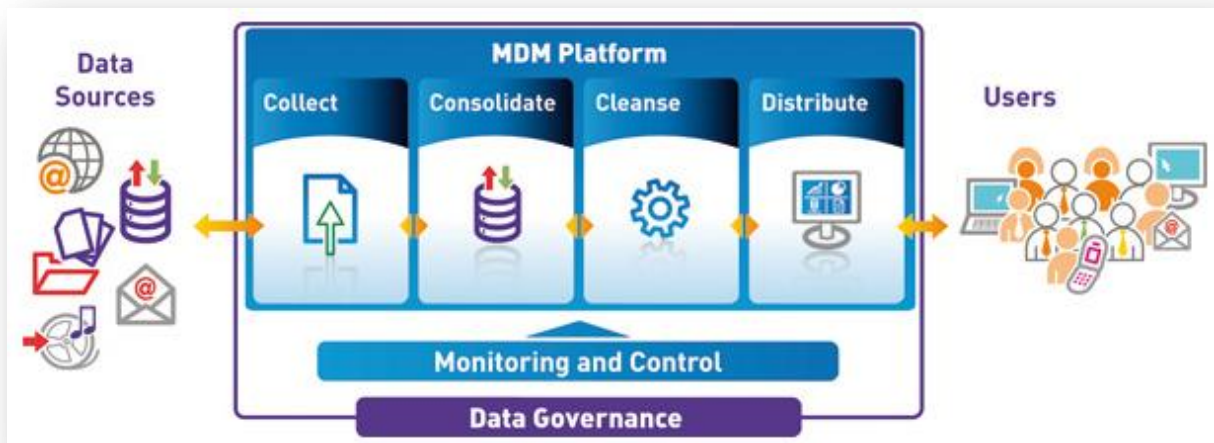
- **L'augmentation du time to market**
Mise sur le marché de nouvelles offres ;
Rapidité de décision grâce à une information et des supports d'analyse fiables ;
Suivi client efficace et réponse client rapide.

- **La réduction du risque d'erreur**
Contrôle et qualité de l'information (data management) ;
Unicité et fiabilité de l'information (master data) ;
Automatisation de process manuels et administratifs ;
Sécurité de la base de données (accès réservé à certaines informations, prévenir des modifications non autorisées).

- **La réduction des coûts**
Rationalisation vis-à-vis des fournisseurs d'information externes ;
Automatisation de process manuels et administratifs ;
Rationalisation de systèmes d'information multiples et complexes.

Typiquement, les besoins conduisant à l'implémentation d'un outil de Master Data Management, sont les suivants :

- Différencier les informations stratégiques et les informations métier, et avoir une vision globale de ces informations stratégiques ;
- Fiabiliser l'information ;
- Bien diffuser l'information (quelle information ? A qui ? Quand ? Où ?) ;
- Continuité historique de l'information (source unique de référence, unicité d'identifiant pour l'ensemble de l'entreprise, traçabilité tant réglementaire que pour des besoins de reporting).



MDM Platform / Source : <http://fr.exalead.com/>

La gouvernance des données est un cadre de contrôle qualité visant à évaluer, à gérer, à exploiter, à optimiser, à contrôler, à entretenir et à protéger les données des entreprises. De nombreux facteurs poussent actuellement la gouvernance des données en tête des préoccupations des décideurs. Il est généralement admis que les sociétés disposant d'un plan efficace dans ce domaine sont capables de produire durablement des données plus précises, complètes et cohérentes sur leurs activités dans l'ensemble de leurs services.

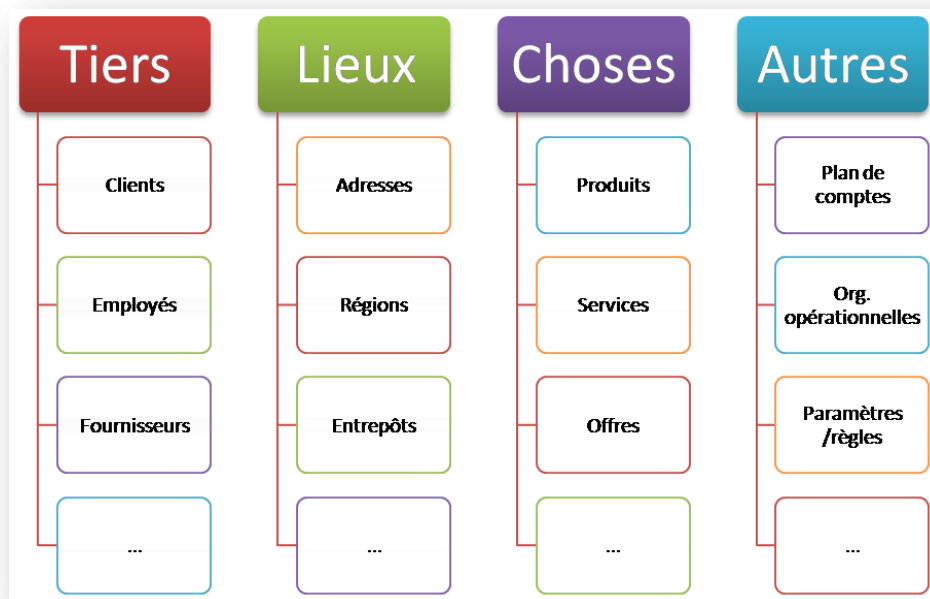
La consolidation de données consiste à rassembler plusieurs données semblables afin d'obtenir un rapport structuré, plus facile à consulter que l'information brute présente sur le serveur, mais avec le moins de perte d'information possible.

Le nettoyage des données permet de vérifier et de corriger les données tout en assurant la standardisation des formats des bases de données des entreprises. Les procédés de nettoyage de données permettent d'optimiser la qualité de l'ensemble des données des entreprises.

Avant d'utiliser une information et de peut-être prendre une décision, il est important de se poser au préalable quelques questions : quelle est cette information ? D'où provient-elle (base de données, table) ? Comment est-elle calculée ? Quand et comment est-elle rafraîchie ? Quelles sont les précautions d'usages ?

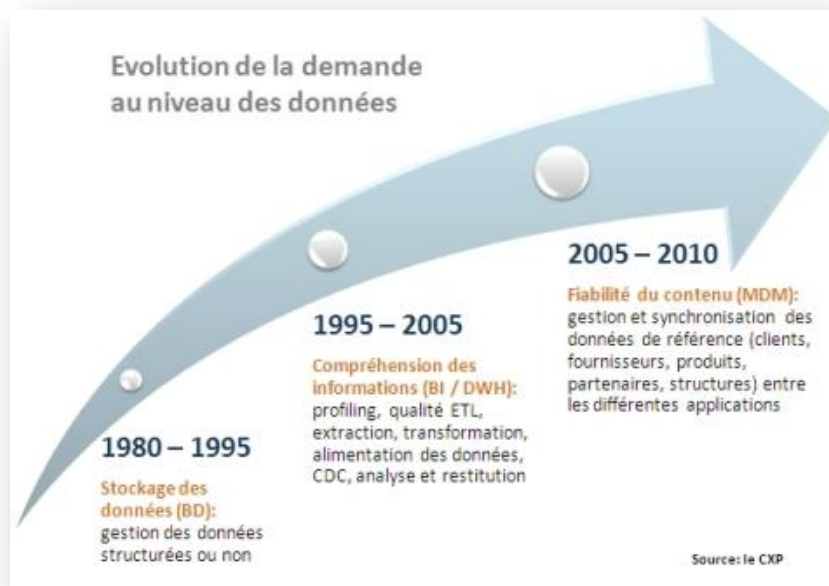
Toutes les informations sur la donnée susceptibles de répondre à cette liste de questions sont enregistrées au sein d'un référentiel. On les appelle **les métadonnées**, c'est-à-dire les données sur les données. Les métadonnées sont actuellement un des points les plus importants de la gestion des systèmes d'information. Une gestion centralisée des métadonnées permettrait de suivre en termes de qualité la donnée à toutes les étapes de sa vie : stockage, extraction, traitement et présentation.

Une **donnée de référence**, ou « master data » est une information de base, unique, fondamentale pour l'activité de l'entreprise, et partagée dans différents services. Par exemple, le nom d'un client, le code d'un produit, la référence d'un fournisseur, un numéro de compte... sont des données exploitées, souvent dans des formats différents, par le service des ventes, le suivi de production, la direction comptable, le contrôle de gestion... La donnée de référence, qui est une donnée métier, doit en effet être identifiable et reconnue comme telle partout dans l'entreprise, quel que soit le service qui en est responsable, quel que soit le système d'information, le serveur ou le logiciel qui l'héberge, la traite ou l'enregistre, ou quelle que soit la division ou la filiale de l'entreprise qui la produit.



Exemples de données référentielles / Source : <http://blog.homsys.com>

D'après l'évolution de la demande au niveau des données : l'entreprise s'est d'abord focalisée sur le stockage des données (avec les systèmes de gestion de bases de données) puis sur la compréhension des données (avec les outils décisionnels dits de BI), et enfin, sur la fiabilité du contenu (avec les outils de MDM).



Source : http://www.cxp.fr/domaine-expertise_mdm.htm

Si la notion de référentiel de données n'est pas nouvelle, le concept-même de MDM est apparu en 2003 et prend vraiment son essor actuellement. Historiquement, celui-ci s'est développé sous deux angles :

- **La gestion des catalogues produits** (Product Information Management) notamment dans les domaines de la grande distribution (retail) et du manufacturing,

- **L'intégration des données clients** (Customer Data Integration) particulièrement pour l'administration de grosses bases de données transactionnelles (gestion des doublons, vérification et homogénéisation des adresses, etc...).

Depuis, ce concept s'est généralisé à d'autres problématiques et concerne la gestion des tiers et des personnes, des produits et des offres, de l'organisation et des structures, des nomenclatures et des codifications, de la configuration et des paramètres.

Les approches sont connues et pratiquées dans l'Enterprise Application Integration (EAI) depuis vingt ans. La terminologie anglophone propre au MDM est issue du travail de David Loshin.

- **Approche dite « Identity Hub »**
Dans cette approche, toutes les données-maître sont gérées dans un registre unique. Ce sont les composants de ce registre (le « Hub ») qui transforment la donnée. Enfin, les applications clientes interrogent le registre. Cette approche a un inconvénient : les systèmes opérationnels ne sont pas prévenus des changements dans le registre.
- **Approche dite « Proactive Harmonization »**
Un seul changement par rapport à l'approche précédente : les systèmes opérationnels sont prévenus des changements. Cette approche et la précédente ont encore un point faible : chaque système opérationnel reste indépendant et les systèmes ne seront pas au même niveau d'intégration.
- **Approche dite « Integrated Hub »**
C'est la « cousine » de l'approche « Integrated ETL » des systèmes d'Intelligence d'affaires développée au début des années 90. Chaque système opérationnel est intégré dans un « MDM Hub » : en clair, ce Hub se substitue à des composants fournisseurs des applications.

Généralement, au sein de l'entreprise, le consensus porte sur la deuxième approche. La première est trop faible : elle oblige les systèmes opérationnels à détecter les changements dans le registre de données-maître. Et la troisième a un coût plus élevé. Dans la vraie vie, plusieurs approches sont utilisées au sein d'une même compagnie. Exemple : un « Integrated Hub » (très réactif) pour les données-maître critiques (client & produit, par exemple) et une « Proactive Harmonization » pour les données-maître restantes.

Normes et standards

Le **Common Warehouse Metamodel** ou **CWM** est un modèle normalisé proposé par l'OMG⁷ qui décrit un langage d'échange de métadonnées à travers un entrepôt de données, un système décisionnel, un système d'ingénierie des connaissances (gestion des connaissances), ou des technologies de portail. Il est fondé sur UML⁸ et XML⁹. Les principaux éditeurs de business intelligence soutiennent ce modèle.

Le **Meta-Object Facility** ou **MOF** est un standard de l'OMG s'intéressant à la représentation des métamodèles et leur manipulation.

L'**Open Information Model** ou **OIM** était un standard soutenu par Microsoft mais qui a baissé les bras et décidé de contribuer au modèle commun, le CWM.

⁷ Object Management Group

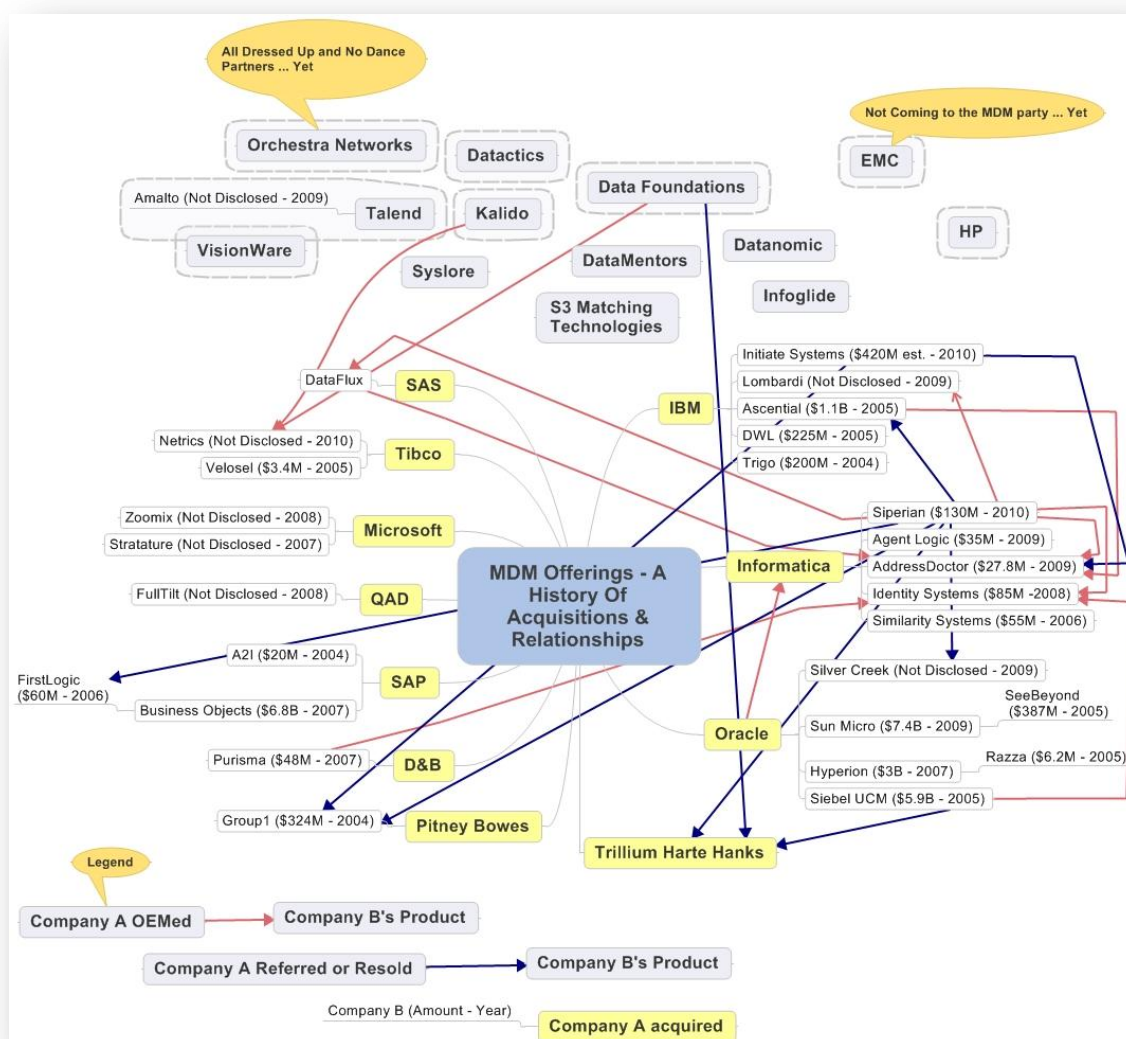
⁸ Unified Modeling Language

⁹ eXtensible Markup Language

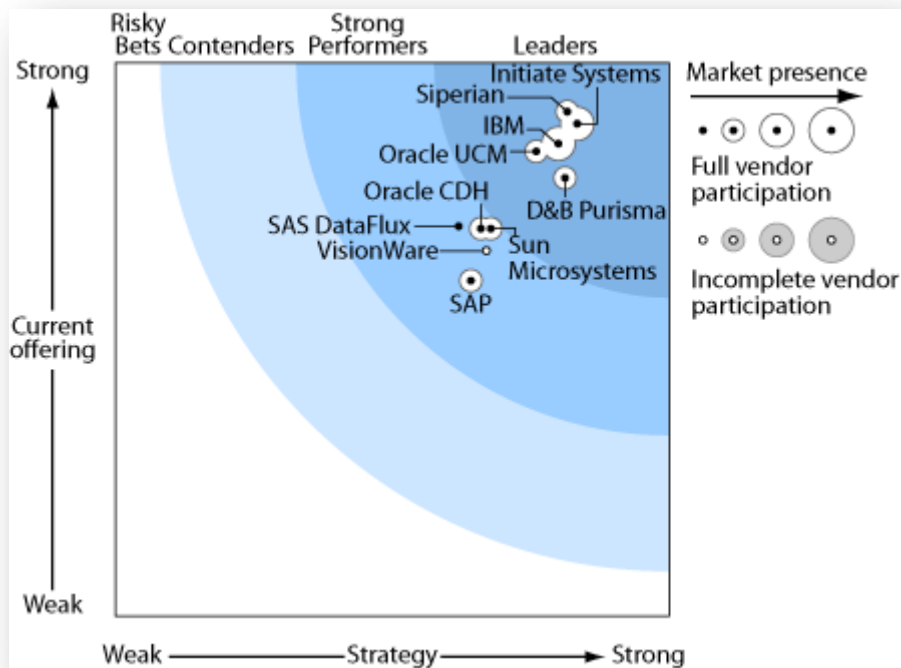
Le marché du MDM

Depuis quelques années, les acteurs de la business intelligence investissent fortement sur ce secteur en pleine consolidation. Le MDM semble faire le « buzz » actuellement et de plus en plus d'entreprises s'y intéressent. Le marché est actuellement alimenté par les fournisseurs de solutions (éditeurs le plus souvent) plus que par les entreprises qui ont besoin de solutions adaptées (case studies). Les fournisseurs se positionnent ainsi en marchand de rêves pour des entreprises qui se retrouvent en position d'enfants dans un magasin de jouets.

Voici une carte intéressante qui retrace les liens et les différents rapprochements entre les acteurs du secteur.



Carte des solutions MDM / <http://www.ramonchen.com/>



Source: August 2008 "The Forrester Wave™: Customer Hubs, Q3 2008"

Il est flagrant que ces acquisitions successives ont fortement impacté des géants comme SAP ou ORACLE, clairement en perte de vitesse et sans réelle stratégie à ce jour.

Les bénéficiaires de ce chamboulement sont les clients et les utilisateurs qui profitent pleinement de l'émulsion du marché pour voir les solutions gagner en maturité.

Les acteurs

Le tableau suivant présente une liste d'éditeurs de MDM et quelques informations sur leur solution.

Leaders

Visionnaires

Pure Player

Spécialistes BI/Enterprise Information Management

Leader Middleware

Perturbateurs

Editeur	Solution et éditeur MDM DQM racheté	Solution BI Complète	Open Source	Informations
IBM	InfoSphere MDM Server WebSphere Initiate Master Data Service Initiate Systems (2010) Lombardi (2009) Ascential (2005) DWL (2005) Trigo (2004)			IBM est l'un des leaders du marché. Il se fait une part belle dans le secteur de la santé puisqu'une quarantaine de services informatiques de santé dans le monde utilisent les solutions d'Initiate Systems. Il existe différentes offres en fonction des données : Websphere Customer Center (gestion données client) et Websphere Product Center (gestion des données produit).
Oracle	Oracle Master Data Management Suite Mural Silver Creek (2009) Sun Micro (2009) Hyperion (2007) Siebel UCM (2005)		X	Talend n'est pas le premier à s'être essayé sur les pentes du MDM Open Source, Sun avait le premier lancé un outil très complet et une communauté de développement Mural pour ce type d'offre.
SAP	NetWeaver MDM IQ8 Integration Studio Business Objects et FirstLogic (2007) A2I (2004)			L'Allemand est l'un des pionniers du Master Data Management. L'éditeur a modifié en 2006 les bases technologiques de son MDM en s'appuyant sur la solution de l'éditeur A2i.
Teradata	MDM Studio MDM Server			Le MDM de Teradata est basé sur le code source de la solution MDM de l'éditeur i2 (partenariat).
Tibco	Collaborative Information Manager Netrics (2010) Velosel (2005)			TIBCO CIM intègre l'intelligence artificielle de Netrics. Le moteur de Netrics fournit un couplage des données temps réel.
Riversand	MDMCenter			Riversand permet aux entreprises de regrouper, gérer et diffuser l'information produit à jour à partir d'un référentiel central.
D&B	Purisma Data Hub Delivers			Purisma est placé parmi les visionnaires selon l'étude du Gartner "Magic Quadrant for Master Data Management of Customer Data" de 2009. Cette offre est par contre quasiment absente du marché français.
Informatica	PowerCenter Siperian (2010) Agent Logic (2009) AddressDoctor (2009) Identity Systems (2008) Similarity Systems (2006)	X		Après avoir réalisé quelques acquisitions dans le domaine de la qualité de données, Informatica a déboursé 130 millions de dollars pour mettre la main sur Siperian, l'un des leaders du MDM. La qualité des données est en quelque sorte un préambule à la mise en place d'une approche MDM.
SAS	DataFlux qMDM DataFlux			SAS lance sa plateforme MDM composée d'un nouveau triptyque : MDM, ETL et DATA QUALITY. Cette nouvelle plateforme baptisée DataFlux Data Management Platform vise à consolider et fiabiliser

				les données de l'entreprise afin d'en tirer le meilleur parti en les transformant en capital stratégique.
Microsoft	SQL Server 2008 Bulldog Zoomix (2008) Stratature (2007)	→	X	SQL Server 2008 R2 propose désormais une fonctionnalité de Master Data Management (MDM) se nommant Master Data Services (Bulldog). Bulldog regroupe les processus et standards Microsoft appliqués au code de Stratature.
QAD	Perfect Product Suite FullTilt (2008)			MDM orienté produit e-commerce intégré à une solution PIM (Product Information Management).
Smartco	Smart Financial DataHub			Pure player français positionné sur le marché du MDM à destination des institutions financières. Outil de gestion de la qualité évolué. Limité en volumétrie traité. Principale référence du marché français.
Talend	Talend MDM Xtensis MDM Amalto (2009)	→	X	La solution Talend MDM Open Source est bâtie autour du noyau de la solution Amalto Xtensis.
VisionWare	MultiVue Identification Server			Visionware continue d'affronter la concurrence dans le segment de marché Microsoft .NET, et a réussi à se développer en Amérique du Nord dans le secteur public et dans la santé en développant une offre moins cher que ses concurrents.
Kalido	Kalido MDM			Kalido MDM est une application logicielle permettant d'harmoniser, stocker et gérer les données de référence au cours du temps.
Orchestra Networks	EBX.Platform			Pure player français positionné sur le marché du MDM. Parmi les nouveautés de la version 4.7RC de EBX.Platform, l'arrivée officielle de la gestion des workflows.
Data Foundations	OneData			OneData utilise un navigateur pour la gestion des données qui permet de consolider, de nettoyer, de créer, de gérer et de normaliser toutes les données partageables, y compris les données de base, les données de référence et les métadonnées.
Health Market Science	CompleteView			Health Market Science propose l'une des offres MDM les plus complètes sur le marché de la santé aux Etats Unis.
Amdocs	Enterprise Customer Hub			Spécialiste des industries de télécommunication. Enterprise Customer Hub est basé sur une version OEM de IBM InfoSphere MDM Server.
Information Builders iWay Software	iWay MDC			iWay Master Data Center (MDC), solution MDM de type CDI, spécialisée sur la gestion des données Clients. iWay MDC permet aux entreprises de consolider des millions d'enregistrements à partir d'un certain nombre de règles métier très faciles à définir.
GXS PMDM	Product Information Manager			Product Information Manager est une solution web qui permet de gérer et d'uniformiser les informations relatives au produit dans l'entreprise.
EnterWorks	Enable MDM			Enable MDM dispose d'une technologie de workflow qui permet de rationaliser les processus liés à un produit et de coordonner les personnes et les systèmes.
Stibo Systems	STEP			STEP est conçu pour fournir un ensemble d'outils et une interface utilisateur intuitive pour la gestion, la validation, la manipulation et l'accroissement des données.
Heiler Software	Product Manager			Product Manager est une plate-forme standard pour la gestion de l'information produit et de la communication produit dans le commerce et l'industrie.

Innovit	ilCE Express Edition (EE)			ilCE Express Edition est une solution MDM complète créée pour les petites et moyennes entreprises qui recherchent un système pour gérer leurs données de produit.
Liaison Technologies	Liaison MDM			Liaison est une plateforme MDM en mode SaaS.
Zycus	Zynapse Automated Master Data Management Solution Suite			Zynapse offre une solution basée sur l'intelligence artificielle entièrement automatisé pour nettoyer, enrichir et dé-dupliquer (éliminer les redondances) les données de référence.
Cegedim dendrite	Nucleus360 as a service			Le premier MDM spécifique aux sciences humaines et disponible en SaaS. Cegedim a pour client un grand groupe pharmaceutique.
Exalead	CloudView			CloudView peut permettre d'accélérer la phase de MDM ou venir remplacer complètement un MDM pour des petites organisations avec des sources et des volumes de données limitées.
Global Ids	ID Integrator			ID Integrator automatise l'intégration de données de référence dans l'entreprise.

On remarque ici que les leaders du marché sont les éditeurs qui ont investi très tôt dans le MDM. On peut prendre l'exemple d'IBM avec le rachat de Trigo Technologies dès 2004. Depuis le rachat de Sun par Oracle, l'idée d'un MDM Open Source basé sur le projet Mural s'est estompée. Talend MDM est donc la seule offre Open Source du marché. A la manière d'Informatica, l'acquisition d'éditeurs d'ILM (gestion du cycle de vie de l'information) ou de spécialistes de la qualité des données peut être un préambule à la mise en place d'une approche MDM.

Le projet MDM

Les points qui vont être énoncés ci-dessous sont considérés comme primordiaux dans le succès d'un projet de MDM.

La gestion des données de référence repose sur plusieurs composantes :

1. Une définition commune et unique de l'information de référence, partagée dans toute l'entreprise (entre les métiers, entre maîtrises d'ouvrage et maîtrises d'œuvre...). Il s'avère souvent complexe de s'accorder sur une définition commune d'une même appellation.
2. Des processus de mise à jour et de validation des données explicitement formalisés et répondant aux exigences de qualité de tous les acteurs de l'entreprise concernés.
3. Des « réceptacles » robustes et pérennes pour les systèmes référentiels de l'entreprise.
4. Des données de référence qui constituent le contenu ; un contenu qualifié par ses « métadonnées » permettant de l'utiliser au mieux : date de mise à jour, de péremption, auteur de la mise à jour...
5. Des applications clientes à adapter pour se connecter au référentiel et utiliser absolument les Master Data.
6. Une roadmap de mise en œuvre du référentiel reste indispensable :
 - Quel est l'état des données existantes (audit des données) ?
 - Comment migrer ces données existantes ?
 - Comment faire évoluer les applications clientes existantes ?
 - Comment maîtriser la connexion au référentiel des nouvelles applications ?
 - Comment gérer le changement ?Autant de questions majeures pour la réussite du projet MDM.
7. Une dimension clé : la mise en place d'une gouvernance des données de référence. À commencer par l'identification d'une « réelle » maîtrise d'ouvrage, responsable du référentiel et des politiques de gestion de son évolution.

D'après Pascal Anthoine, directeur conseil chez Micropole-Univers, « le budget des projets MDM n'est pas si dissuasif si on le compare aux gains aussi bien en termes de qualité de donnée que d'optimisation des processus. Avec le recul sur la vingtaine de projets MDM menés chez Micropole-Univers, le budget moyen tourne plus autour des 300 à 400 k€ que du million d'euros, comme ont pu l'annoncer certains acteurs... ».

Les projets de MDM ont des impacts majeurs en termes d'outils, d'application, d'organisation et de processus. Il s'agit donc de projet long terme dont les bénéfices, stratégiques pour l'entreprise, ne sont accessibles qu'après une longue traversée du désert.

Perspectives et tendances

Nous entrons dans une nouvelle décennie et sept nouvelles tendances se dégagent autour du nouveau monde du MDM :

1. Les clients ne veulent plus d'applications horizontales (progiciel conçu pour être utilisé par différents types d'organisation). Le MDM doit s'adapter aux exigences spécifiques des industries. Les résultats doivent être pertinents à la façon dont une industrie fonctionne.
2. Les clients souhaitent avoir des outils pour lier des hiérarchies et des relations sur les données non structurées afin de connaître la valeur de l'information.
3. L'utilisation des modèles Cloud Computing et SaaS vont imposer les approches hybrides. Les systèmes MDM doivent prendre en charge les modèles hybrides en temps réel. L'intégration des données doit être pilotée par les événements.
4. La manière de manager les données de référence n'a pas d'importance, seuls les résultats comptent. Les utilisateurs de BI recherchent des résultats et des pistes de réflexion.
5. La gouvernance et la gestion des données deviennent les points les plus importants du MDM.
6. Il est nécessaire de créer une relation de confiance, de devenir un partenaire, un vrai, plutôt qu'un fournisseur froid et distant, si l'on souhaite justement gagner la fameuse confiance tant recherchée des consommateurs. Les organisations ont besoin de comprendre aussi bien leur défenseur que leur détracteur. Dans un monde social et interconnecté, il est nécessaire d'avoir un marketing, des ventes et un support plus ciblés.
7. Le MDM va devoir couvrir plus de types de données. De nouvelles formes de contenu entreront bientôt dans l'équation (images, vidéos, flux, tweet...).

IBM Research a annoncé en janvier 2008 le lancement du projet Semantic Master Data Management (SMDM) qui vise à utiliser les technologies Web sémantique pour outiller l'interrogation et l'analyse sémantique de données de référence. Cette technologie vise à terme à améliorer l'utilisation des données de référence en se greffant aux solutions MDM existantes.

Le projet SMDM a fait le choix de gérer des ontologies enrichies à partir du modèle de données logique d'une solution MDM. Pour mémoire, l'ontologie est un concept fondamental du Web sémantique, pouvant être définie comme la spécification explicite d'une conceptualisation partagée. Il s'agit donc d'une structure clé pour la représentation des connaissances. Les données maîtres sont représentées sous formes d'instances d'ontologies de la forme < sujet, prédicat, objet >, similaires au langage naturel et aisément interprétables. Après la mise en place de l'entrepôt d'ontologies, celle des moteurs de recherche associés : le langage de requête SPARQL permet l'interrogation de ces données, les requêtes étant intuitivement construites sans qu'il ne soit nécessaire d'avoir connaissance du modèle de données sous-jacent. L'utilisation d'ontologies permet également la découverte d'informations implicites par le biais du raisonnement ontologique : découvertes de nouvelles relations, classification automatique, mise en relation automatique, catégorisation à la volée.

Les solutions MDM et de gestion de contenu actuelles maintiennent elles-mêmes leurs propres métadonnées, et de ce fait l'entreprise reste tributaire des choix (trop souvent propriétaires) effectués par chaque éditeur. Ces silos de métadonnées sont un obstacle pour l'entreprise à l'accès à une compréhension globale et transverse de ses données stratégiques, structurées ou non structurées.

Par la construction d'un socle sémantique utilisant les standards W3C au dessus de ses entrepôts de données, l'entreprise se dote des moyens lui offrant de nouveaux usages pour l'exploitation des contenus métiers, par la capacité à valoriser l'ensemble de ses données stratégiques en véritables informations. Alors que la sémantique permet la compréhension, la déduction automatique et l'interrogation intelligente des données, les standards sémantiques ajoutent l'interopérabilité inter-applicative : les outils d'analyse ont accès à toujours plus de contenu intelligible ; les règles métiers peuvent s'affiner et perfectionner la gouvernance des données, la déduction automatique permet la mise en relation de données toujours plus hétérogènes, et l'interrogation sur ces données devient plus pertinente, offrant ainsi une meilleure maîtrise des actifs informationnels de l'entreprise.

Conclusion

De nos jours, la business intelligence évolue sans cesse, ce qui nécessite une attention toute particulière pour se tenir informé des dernières tendances. De plus, la BI est un domaine où les enjeux sont considérables et pour lequel il est très important d'anticiper sur l'avenir.

Nous avons donc veillé sur différents axes technologiques et avons tenté de pressentir ce qui pourrait être les futures tendances dans le monde du décisionnel. Deux axes ont particulièrement retenu notre attention, le Master Data Management et les nouveaux médias de restitution, et ce pour diverses raisons. Le marché du MDM est aujourd'hui en pleine expansion et la qualité des données est un élément primordial en BI. C'est pourquoi le MDM est le premier axe sur lequel nous pourrions investir. Viennent ensuite les médias de restitution qui continuent à évoluer. Il serait bénéfique d'investir dans ce domaine car des médias de restitution originaux et performants peuvent être des atouts de poids pour conquérir de nouveaux clients.

Sources

Les ETL

<http://grim.developpez.com/articles/concepts/etl/>

<http://www.decideo.fr/>

<http://www.journaldunet.com/solutions/0702/070221-panorama-etl/1.shtml>

Les nouveaux modes de stockage

In memory

http://en.wikipedia.org/wiki/Flash_memory

Qlikview

<http://www.pvirgule.fr/contenu-communique/3202/429/uk?pv=3a9ea1715e1dd24a6f96a302086d1fd6>

<http://www.qlikview.com/fr/explore/products/in-memory-advantage>

http://www.decideo.fr/Edition-2010-du-barometre-Decideo-QlikView-remporte-le-Decideo-d-Or_a3527.html

SAP

<http://www.sap.com/france/solutions/index.epx>

<http://searchbusinessanalytics.techtarget.com/news/1507060/In-memory-technology-promises-faster-more-flexible-BI-and-data-analysis>

<http://www.lemagit.fr/article/serveurs-lames-sap-decisionnel-oracle-sybase-sgbd-infotel-erp-dram-lames-bydesign-snabe-business-suite-in-memory-plattner-blade/5221/1/in-memory-comment-sap-compte-ringardiser-les-bases-donnees-surtout-marginaliser-oracle/>

La business intelligence en mode SaaS

Rapport du cabinet Hurwitz & Associates- 'The TCO Advantages of SaaS-Based Budgeting, Forecasting & Reporting'-2010

Rapport du Gartner - '*Business Intelligence as a Service: Findings and Recommendations*' - 2009

Rapport du Gartner - '*Hype Cycle for Software as a Service*' - 2009

Rapport du Gartner - '*Key Issues for Software as a Service*' - 2010

<http://www.channelnews.fr/expertises/tendances/6479-leconomie-du-saas-gagne-en-densite-et-en-maturite-.html>

Marché et acteurs

Les éditeurs et leurs solutions

<http://www.legrandbi.com/2009/03/bi-saas-controverse-sur-l%E2%80%99opportunit%C3%A9-du-modele/>

<http://www.legrandbi.com/2010/02/sap-relance-saas-businessobjects/>

Les cibles

<http://biguru.wordpress.com/2008/03/13/saas-bi-software-as-a-service-model-in-business-intelligence/>

<http://searchbusinessanalytics.techtarget.com/news/1506987/SaaS-BI-and-data-warehousing-finally-stepping-out-of-the-shadows>

<http://searchbusinessanalytics.techtarget.com/news/1506980/SAP-unveils-SaaS-business-intelligence-suite-for-the-rest-of-us>

Changements apportés par le Saas

Pour les éditeurs/intégrateurs

<http://www.legrandbi.com/2009/03/bi-saas-controverse-sur-l%E2%80%99opportunit  -du-modele/>

Pour les clients

http://www.computerworld.com/s/article/9142307/SaaS_BI_vendor_Oco_aims_to_help_enterprise_IT_be_tactical

<http://applicationdevelopment.searchoracle.com/document;5143880/oracle-research.htm>

<http://www.itworld.com/saas/62953/gartner-dont-assume-saas-cheaper>

<http://www.b-eye-network.com/view/13726>

Stockage de donn  es et s  curit  

http://www.decideo.fr/Le-point-sur-les-offres-dites-SaaS-Software-as-a-Service,-ASP-Application-Services-Provider-ou-OnDemand_a3184.html

La business intelligence mobile

Debra Logan, Bill Hostmann, Enterprise Information Management: A Requirement for Enterprise-Scale Business Intelligence and Performance Management Initiatives, Gartner, 2009

Acteurs

<http://www.bee--eye.com/>

<http://www-01.ibm.com/software/data/cognos/products/cognos-8-go/mobile/features-and-benefits.html?>

<http://www.novediagroup.com/>

<http://www.pentaho.com/launch/iPhone/>

<http://www.prelytis.com/>

<http://www.qlikview.com/>

<http://regard-solutions.blogspot.com/>

<http://www.roambi.com/>

<http://www.sap.com/solutions/sapbusinessobjects/business-intelligence/information-infrastructure/enterprise/mobile/index.epx>

<http://www.webalo.com/>

March   des smartphones

<http://www.canalys.com/pr/2010/r2010021.html>

<http://france.blog.nimbuzz.com/2009/10/16/les-mobiles-sous-android-plus-repandus-que-les-iphone-dici-a-2012-mais-symbian-toujours-en-tete/#more-540>

S  curit  

http://www.decideo.fr/CINQ-MESURES-CLES-POUR-PROTEGER-LES-DONNEES-DECISIONNELLES-CONSULTABLES-SUR-LES-TERMINAUX-MOBILES_a3291.html

<http://www.indexel.net/securite/smartphones-cinq-regles-de-securite-a-respecter.html>

Apports de la mobilit  

<http://www.usinenouvelle.com/article/l-avenement-du-decisionnel-mobile.N127147>

<http://www.idexen.com/blog/cognos-8-go-mobile-est-disponible-sous-windows-mobile/>

Les moteurs de recherche

Introduction

Article Wikipedia sur les moteurs de recherche. http://fr.wikipedia.org/wiki/Moteur_de_recherche

L'importance du moteur de recherche

Source : 12 moteurs de recherche d'entreprise comparés par Denis Lafont, responsable du pôle de compétences Enterprise Search de Capgemini. <http://www.eweekurope.fr/produit/comparatif-12-moteurs-de-recherche-d-entreprise-au-crible-745>

Le marché et les acteurs

Deux pages PDF sur l'entreprise Sinequa. <http://www.evalir.com/resources/sinequa-connect-to-Knowledge.pdf>

Les technologies du web sémantique

Les informations non-structurées.

http://benhur.telug.uqam.ca/SPIP/inf6104/article.php3?id_article=17&id_rubrique=4&sem=2

Certains articles du blog <http://nicolas.cynober.fr/>, co-fondateur de pearltrees. Lien vers le pearltrees Web sémantique : http://www.pearltrees.com/#/N-u=1_2&N-f=1_18&N-s=1_18&N-p=82&N-play=0

<http://websemantique.org/>

<http://wiki.dbpedia.org/>

PDF du Gartner « Hype Cycle for Enterprise Information Management, 2009 » page 8

Pôle innovation d'Exalead. <http://fr.exalead.com/software/company/innovation/>

Page d'actualités de Sinequa. <http://www.sinequa.com/fr/news,l-enterprise-search-devient-une-application-decisive-avec-la-sortie-de-la-version-7-0-de-sinequa-enterprise-search,188.html>

Une interview de Fabrice Lacoix, président d'Antidot. <http://www.clubic.com/actualite-79037-fabrice-lacroix-antidot-web-semantique.html>

Blog de l'éditeur Antidot. <http://blog.antidot.net/tag/web-semantique/>

Le futur des moteurs de recherche

Enterprise Search Bus. <http://jean-ferre-blog.blogspot.com/2010/02/entreprise-search-bus-intelligent.html>

Produits Exalead On Demand. <http://www.exalead.com/software/products/on-demand/>

Les médias de restitution

Indexation

http://www.photogestion.com/manuel/indexation_image.html

<http://www.educnet.education.fr/dossier/metadata/metadonnees>

IMEDIA

<http://www.inria.fr/recherche/equipes/imedia.fr.html>

<http://www.heritagemuseum.org/fcgi-bin/db2www/qbicSearch.mac/qbic?sellLang=English>

CQBI

<http://www.qbic.almaden.ibm.com/>

<http://spiedl.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=PSISDG001908000001000173000001&idtype=cvips&gifs=yes&ref=no>

Gaudi

<http://www.clubic.com/actualite-164322-google-audio-indexing.html>

<http://labs.google.com/gaudi>

Aigle

<http://www.business-geografic.com/company/>

Business Geographic sur Iphone :

<http://www.directionsmag.com/press.releases/?duty=Show&id=48984>

Exalead

Exalead devient partenaire de Micropole-Univers :

http://www.cfo-news.com/Exalead-devient-partenaire-de-Micropole-Univers_a14387.html

Quaero

<http://www.quaero.org/media/files/presse/fr-brochure-all.pdf>

LTU

<http://www.ltutech.com/fr/press-articles/ltu-lance-une-api-de-recherche-visuelle>

<http://www.ltutech.com/fr/technology/matching-de-videos>

<http://www.publi-news.fr/data/28042010/28042010-093715.html>

<http://www.itchannel.info/articles/104150/ltu-lance-api-recherche-visuelle.html?key=ca3f96286fb5f4d3>

<http://www.lookthatup.com/>

Chromatic

<http://www.batiproduitsmaison.com/bpm/FicheProdRecherche2.do?id=1475385958>

Autonomy

<http://www.autonomy.com/>

Google Location Services

<http://fr.www.mozilla.com/fr/firefox/geolocation/#geo-demo>

NewPhenix

http://societe.journaldunet.com/fiche/chiffre_cle/24964/newphenix.shtml

MDM

Site d'informations et d'actualités sur la Business Intelligence

http://www.decideo.fr/Des-solutions-de-gestion-des-donnees-de-reference-MDM_a3293.html

http://www.decideo.fr/2010-l-annee-du-Master-Data-Management_a3592.html

http://www.decideo.fr/Gerer-le-Referentiel-Metier-dans-l-Entrepot-de-Donnees_a2853.html

<http://www.legrandbi.com/2010/01/informatica-siperian-mdm/>

<http://www.piloter.org/business-intelligence/projet-mdm.htm>

Site d'actualités générale et business

<http://pro.01net.com/>

<http://www.journaldunet.com/solutions/acteurs/strategie-mdm-de-talend.shtml>

Site d'éditeur ou de cabinet de conseil

<http://www.gartner.com/technology/media-products/reprints/oracle/article78/article78.html>

<http://www.dataflux.com/dfblog/?p=2032>

<http://www.dataflux.com/dfblog/?tag=master-data-management>

<http://www.teradata.com/master-data-management>

http://www.cxp.fr/domaine-expertise_mdm.htm

http://www.informationdifference.com/mdm_landscape.html

Blog business et MDM

http://nourdinebouaghaz.blogspot.com/2010/02/une-image-vaut-mieux-que-mille-discours_09.html

http://nourdinebouaghaz.blogspot.com/2010/02/ibm-degaine-son-tour-et-s-initiate_07.html

<http://nourdinebouaghaz.wordpress.com/2010/03/07/mais-au-fait-le-mdm-cest-quoi/>

http://nourdinebouaghaz.blogspot.com/2009/12/reflexion-mdm-les-entreprises-sont_3583.html

<http://nourdinebouaghaz.blogspot.com/2010/04/le-chaos-ruineux-de-la-non-qualite-des.html>

<http://blog.homsys.com/tag/mdm>

<http://architecturedinformation.blogspot.com/2010/03/mdm-les-approches.html>

<http://architecturedinformation.blogspot.com/2010/03/master-data-management.html>

http://blogs.forrester.com/business_process/2010/01/introducing-the-mdm-markets-newest-800lb-gorilla-informatica-acquires-siperian.html

<http://blog.homsys.com/tag/mdm>

<http://www.masterdatamanagement.fr/master-data-management>

<http://oliviermathurin.blogspot.com/2008/07/perspectives-mdm-et-technologies-web.html>

Livre blanc et livre

SearchDataManagement.com's Master Data Management Platform Directory 2009 Edition

La gestion des données de référence avec TIBCO Collaborative Information Manager

La gestion des données de référence ou comment exploiter toutes vos informations (IBM)

Master Data Management and Data Quality Dataactics White Paper

Les nouveaux tableaux de bord des managers Alain Fernandez