Livre blanc

Veille technologique sur la Business Intelligence Temps Réel

Mai 2012

POLYTECH LYON

BENSAHLA Adel

BERARD Aurélie

LUZY Yannick

MURE Mathieu

PERRAUD Alexandre

VILLAGE Benoit

Remerciements

Nous tenons à remercier tout d'abord l'ensemble de la société XIA, qui a commandité ce projet. Nous tenons tout d'abord à remercier Mr DOSSOU qui a été notre premier interlocuteur et qui a su poser les attentes de son entreprise par rapport à la veille sur la business intelligence temps réel. Puis à Mr DLIOUAH Abdallah et Mr FOURNIER Jean-François qui ont su nous guider et conseiller sur la suite de la veille technologique.

Nous remercions également Mr BONNEVAY Stéphane et Mr LALLIARD Michel nos responsables pédagogiques en qui ont su répondre à nos questions et qui ont suivi cette veille technologique avec beaucoup d'intérêt et de motivations.

Nous remercions également Mr MOHAND ACID Saïd et Mr ELGHAZEL Haytham, enseignants chercheurs au sein de l'université Lyon 1, experts dans le domaine de la business intelligence qui ont su nous orienter et nous présenter le monde de la business intelligence ainsi que des pistes qui se sont révélées très instructives et très intéressantes.

Nous remercions finalement tout le département informatique de Polytech Lyon sans qui cette veille technologique n'aurait pu être possible.

Sommaire

Introduction

Dans le cadre de notre formation au sein de Polytech Lyon, nous participons à une unité d'enseignement de veille technologique.

Notre groupe est composé de six étudiants et encadré par deux responsables pédagogiques, ce module consiste, pendant six mois, à veiller sur un domaine technique à partir d'une problématique proposée par un commanditaire.

Ce dernier a souhaité connaitre les évolutions et les technologiques existantes concernant la business intelligence temps réel, technologie en actuel développement.

Ce sujet de veille s'inscrit dans une optique de mise à jour du système décisionnel au niveau du traitement des informations concernant les données des ressources humaines dans l'entreprise de notre commanditaire.

Dans un premier temps, nous avons recherché des informations sur la Buisiness Intelligence Temps Réel. Nous avons pu ainsi acquérir des connaissances sur le sujet et ainsi effectués un état des lieux dans ce domaine.

Nous avons ainsi dégagé neuf axes principaux nous permettant de répondre à la problématique posée sur lesquels nous basés nos recherches :

- Cloud BI
- Compression des données
- Techniques de sécurisation des données
- Solutions matérielles
- Techniques d'exploration et de restitution des données
- Le datamining en temps réel (accès concurrents et performance)
- L'alerte temps réel
- Les libertés informatiques
- BI mobile
- Les restrictions législatives concernant le trading boursier

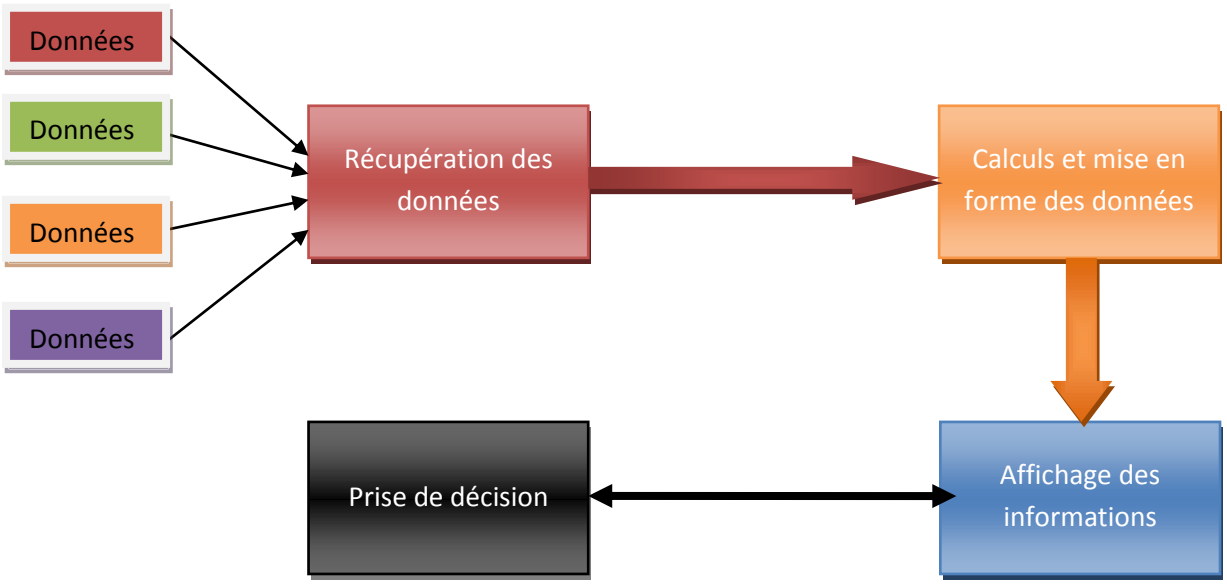
Ce livre blanc reflète le travail effectué au cours de cette période de veille, il expose les résultats de nos recherches, et conclu en envisageant le développement futur de ce secteur.

Partie 1 : Pré-requis

I. Généralités

A) La Business Intelligence

L'informatique décisionnelle ou Business Intelligence représente les outils ainsi que les méthodes permettant de récupérer, de transformer, de calculer et de fournir les données d'une entreprise. Elle permet donc d'offrir une aide à la décision et, par conséquent, permet d'établir la stratégie d'entreprise car elle fournit l'ensemble des informations de l'activité.
La Business Intelligence a donc pour but de permettre la meilleure décision possible à un instant donné.
De manière générale, l'informatique de décision suit le schéma suivant :



1. Tout d'abord, il faut récupérer des données hétérogènes qui proviennent de sources différentes.
2. Puis une fois ces données acquises, un traitement (somme, moyenne, …) est appliqué. Ces données sont ensuite formatées et rangées de façon optimale pour répondre aux besoins des futures décisions.
3. Ensuite des informations en sont extraites comme les règles d'associations ou encore des calculs de prévisions.
4. A partir de ces informations, une décision peut être prise.

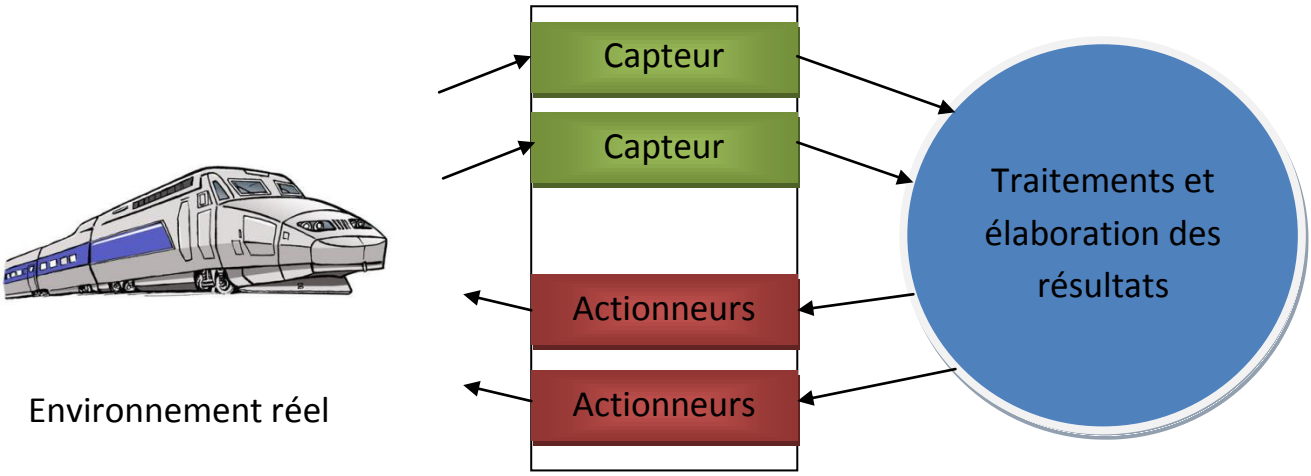
L'informatique décisionnelle apporte donc de l'information élaborée qui interagit directement sur ses stratégies internes. Elle permet ainsi de :

- Comprendre, de gérer et de maintenir sa compétitivité
- Accroître sa part de marché
- Fidéliser sa clientèle
- Optimiser ses processus et ses coûts.

B) Les systèmes temps réel

Le fonctionnement général des applications temps réel peut se résumer ainsi :

- Acquisition des données depuis l'environnement à l'aide de capteurs
- Traitement des données et élaboration des résultats au bout d'un délai limité
- Envoi d'ordres de commande de l'environnement à l'aide d'actionneurs



Les installations temps réel sont présentes dans de nombreux domaines d'application, comme les usines, l'aéronautique, le transport, la gestion des personnels, le service après-vente, la vente etc. Cependant, toutes les applications temps réel ont en commun la prédominance du facteur temps. En effet, les applications temps réel doivent réagir en tenant compte de l'écoulement du temps, c'est la caractéristique fondamentale qui distingue globalement les applications temps réel des autres types d'applications informatiques.
Dans beaucoup d'applications temps réel, l'objectif est la gestion efficace des données dans une contrainte temporelle imposée. Les systèmes adaptés pour la gestion d'applications de ce type sont les systèmes temps réel (STR), car ils utilisent des algorithmes optimisés d'ordonnancement des tâches de manière à respecter les échéances, ainsi que des canaux de communications spécifiques et qui gèrent les contraintes de temps.
Un STR est donc un système qui doit fournir des données exactes, mais ces données doivent être présentes avec un intervalle de temps donné. En effet, suivant la priorité des données, certaines doivent arriver pendant un laps de temps précis. Si elles arrivent trop tôt elles ne sont pas utilisables car elles ne représentent pas la réalité. Et si elles arrivent trop tard, elles sont jugées obsolètes.
Un STR est conçu de manière à définir les réactions de celui-ci dans tous les cas ainsi que de savoir à l'avance si un système va respecter ses contraintes temporelles.

On distingue trois types de temps réel :

- Temps réel dur (*hard real time*) : si les délais ne sont pas respectés alors le système passe dans une situation critique (centrale nucléaire, avion, …).
- Temps réel ferme (*firm real time*) : si les délais ne sont pas respectés alors le système va réagir en signalant une erreur et provoquer une exception.
- Temps réel mou (*soft real time*) : le retard des délais n'est pas catastrophique.

C) La BI Temps réel

Les STR possèdent des mécanismes pour gérer les contraintes temporelles de leurs tâches mais le volume des données qu'ils fournissent reste limité. La Business Intelligence (BI) temps réel est une méthode d'aide à la décision qui permet de gérer la double contrainte, cohérence logique et cohérence temporelle des données.

La BI Temps réel va donc permettre de prendre des décisions à partir de données fournies par des systèmes temps réel (STR).

II. Historique de la Business Intelligence

Le but de l'informatique décisionnelle est de transformer les données de l'entreprise en 'Intelligence'. Il est important de préciser que le mot 'intelligence' est à prendre dans son acception anglo-saxonne, c'est-à-dire connaissance de l'entreprise ; connaissance de son fonctionnement, de ses partenaires : clients et fournisseurs, de sa structure : groupe et filiale, de ses produits, de ses processus, de son organisation, de ses ressources humaines, de son histoire, de son présent, mais aussi de son futur probable.

Le concept d'entrepôt de données (DataWarehouse) est bien plus ancien que ce que la croyance commune veut. Par un concept trop divers, trop ambigu et carrément déroutant, la route pour y arriver a été rude. Heureusement, nous sommes entrés dans une phase de synthèse dans laquelle un nouveau consensus de la conception a émergé. Nous vivons dans un monde où les décisions ne peuvent plus être prises uniquement sur des coups de génie. Nous pouvons aujourd'hui fournir des informations de qualité aux managers qui en ont besoin pour comprendre le présent et simuler l'avenir. Mais le cheminement n'est pas fini, car nous sommes sur le point de prendre un chemin plus radical : celui de l'entreposage de données en temps réel.

Nous allons détailler dans la suite de cette partie l'histoire de l'informatique décisionnelle (Business Intelligence). Permettant ainsi de discuter quant aux évolutions possibles de celui-ci et son effet sur l'architecture d'une entreprise.

La Business Intelligence telle qu'on la connaît aujourd'hui, a évolué à partir de la recherche opérationnelle qui a commencé dans les années 1960, suivie par la supervision (DSS : Decision Support System) dans les années 1970, l'aide à la décision dans les années 1980 (entrepôts de données, OLAP et BI) et l'exploration de données (data mining) dans les années 1990.

A) Les débuts

Tout commença en 1958 où un chercheur, Hans Peter Luhn, travaillant chez IBM publia un article utilisant pour la première fois le terme Business Intelligence (BI). Il la définit comme « la capacité à appréhender les interrelations des faits présents de manière à orienter l'action vers un but désiré ».
Dans les années 1970, les études du MIT cherchant à développer une architecture technique optimale ont permis une évolution du management. Grâce à la mise en place d'une ressource solide d'information, ils avaient l'intention d'élaborer une nouvelle architecture système en partant de zéro. Un principe de base a émergé : séparer le traitement opérationnel et analytique, dispersant ainsi les données dans plusieurs DataWarehouses. Pour l'époque, c'était un concept radicalement différent de ce qui se faisait : mais c'était difficilement réalisable, du fait des capacités de stockage limitées.

B) Années 1980

Digital Equipment Corporation (DEC) a été la compagnie la plus avancée techniquement au monde. Elle a été la première à construire un réseau numérique permettant d'héberger ses applications distribuées. Elle a été aussi la première à migrer ses bases de données relationnelles vers son propre DBMS (DataBase Management System). Pour se développer, ils ont formé une équipe multidisciplinaire comportant ingénieurs, financiers et responsables du marketing. Ils avaient pour mandat d'inventer une nouvelle architecture, mais aussi de l'appliquer sur des moyens numériques et aux systèmes financiers mondiaux. Partant des recherches du MIT, ils décident d'implémenter plusieurs applications distribuées utilisant des DataWarehouses séparés. C'était la première fois qu'un modèle client/server était testé.

Pendant ce temps, IBM poursuit une lutte contre un aspect différent du problème de gestion de l'information. L'un des plus gros problèmes étant d'intégrer des données issues de nombreux systèmes d'informations distincts possédant chacun des codages différents. En 1998, Barry Devlin et Paul Murphy travaillant dans la filiale irlandaise d'IBM abordent le problème d'intégration différemment. Ils introduisent le terme « information warehouse » pour la première fois défini comme : « Un environnement structuré de soutien aux utilisateurs finaux dans la gestion d'une entreprise ». Ils sont soutenus par « l'Information Technology Department » pour la gestion des données. IBM voit là un billet pour sa croissance. Mais où était l'architecture ? Comment allaient-ils construire une telle chose ?
Un terme, datamining (exploration de données), apparut dans les années 1980 grâce à Rakesh Agrawal (informaticien) lorsqu'il entamait ses recherches sur des bases de données d'un volume de 1 Mb. Le concept d'exploration de données fait son apparition, d'après Pal et Jain, aux conférences de l'IJCAI en 1989. Gregory Piatetsky-Shapiro chercha un nom pour ce nouveau concept dans la fin des années 1980, aux GTE Laboratories. « Datamining » étant sous la protection d'un copyright, il employa l'expression « Knowledge discovery in databases » (KDD). Le datamining peut être considéré comme une partie intelligente de l'informatique décisionnelle et il est certain que c'est souvent un élément démultiplicateur de retour sur investissement.

En 1989, Howard Dresner (par la suite devenu un analyste du Gartner Group) a proposé « Business Inteligence » comme terme générique pour décrire les « concepts et méthodes pour améliorer la prise de décisions par des entreprises utilisant des systèmes utilisant des faits ». Il a fallu attendre la fin des années 1990 pour que cette utilisation soit généralisée.

C) Années 1990

Maintenant que l'architecture a été mise en place, il fallait des évolutions majeures du côté client, de l'interface graphique, un développement objet et la possibilité d'utiliser les ressources à travers les différents réseaux. Le développement d'outils d'extraction, de transformation, de chargement (regroupé par la suite sous le terme ETL), de validation, d'interface graphique à été possible grâce à Apple qui a été le premier à proposer un système d'exploitation utilisant une interface graphique. Le développement de frameworks a alors débuté.
En 1991, Bill Inmon publie son premier livre sur le data warehousing qui pour l'époque donnait la définition la plus large de celle-ci. Il se focalisa exclusivement sur ce dont les entreprises avaient besoin. Il mit en place un guide, sur comment construire un DataWarehouse. La définition du DataWarehouse de Bill Inmon correspond encore à celle utilisée aujourd'hui.

DataWarehouse (définition par Bill Inmon) :

Un DataWarehouse est un orienté sujet, intégré, variant dans le temps, possédant une collection de données non volatiles permettant la prise de décisions.

- Orienté sujets : contrairement aux bases de données opérationnelles généralement orientées processus, les bases des entrepôts décisionnels sont modélisées pour répondre facilement à toutes les questions d'utilisateurs non-informaticiens. Il faut donc apporter à l'utilisateur l'information selon sa définition métier.
- Intégré : regroupant généralement des sources hétérogènes de données. Exemple : pour avoir la connaissance métier d'un client, il faut souvent rassembler les informations issues des systèmes opérationnels
  - De gestion des forces de ventes pour connaître ses dernières commandes
  - De la comptabilité pour savoir s'il a payé sa dernière facture
  - Du service après vente s'il l'a utilisée
  - Du serveur Web s'il s'est connecté sur le site

- Du partenaire avec lequel un programme de fidélité commun a été mis en place
- Etc.

Un entrepôt de données décisionnelles d'entreprise doit permettre d'avoir une vision unique et transversale de l'information.

- Variant dans le temps : c'est à dire datés, afin de conserver un historique. Cela permet les analyses comparatives. Exemple : le solde du compte en banque du client est une variable volatile qui change à chaque transaction. On va donc retrouver généralement, si la volumétrie le permet, l'information de détail, le plus petit dénominateur commun à différents problèmes. Ayant ce détail daté, il sera alors toujours possible de recalculer un indicateur à un instant *t*.

- Non volatile : stables, non modifiables. Contrairement à un système opérationnel modifié après chaque transaction, les informations d'un système décisionnel ne changent quasiment pas.

En 1994 nous voyons apparaitre un découpage des données du DataWarehouse vers plusieurs magasins de données, DataMart en anglais (littéralement magasin de données) désignant un sous-ensemble du DataWarehouse contenant les données du DataWarehouse pour un secteur particulier de l'entreprise (département, direction, service, gamme de produits, etc.). On parle par exemple de DataMart Marketing, DataMart Commercial... Le début des DataMarts a été quelque peu difficile. En effet, il était difficile de vendre et concevoir des systèmes importants de cette solution. Il a fallu commencer petit pour grandir progressivement en commençant par l'implémentation à un service pour en « conquérir » d'autres.
Convaincre les clients de construire petit à petit leur entrepôt a permis de sauver l'industrie dans la BI dans le court terme. Il n'y a aucun doute que ce fut un mal nécessaire. Cependant, cela violait le principe le plus fondamental du DataMart. En effet avec un seul DataMart chaque département nettoie, transforme et modifie les sources de données, ce qui a eu pour conséquence d'augmenter le chaos, alors que les DataMarts étaient conçus pour l'éliminer cette incohérence.
Même avec cette architecture rétrécie, de nombreux projets ont échoué. Certaines raisons sont classiques : défaillance du département IT, manque d'une stratégie d'entreprise claire ou mal suivie.

Ralph Kimball, informaticien et chef d'entreprise américaine, publie son premier livre *The Data Warehouse Toolkit*. Le marché connut un nouveau rebond, tout comme un peu plus tôt avec les DataMarts. Ce best-seller a fourni des orientations de conception détaillée sur la façon d'optimiser les données pour une analyse future. La modélisation dimensionnelle enjambe le fossé de la conception relationnelle traditionnelle aux bases de données multidimensionnelles qui a donné naissance au surnom OLAP (Online Analytical Processing) et toutes les variantes que nous connaissons (R-OLAP, H-OLAP, S-OLAP)

Une nouvelle technologie fait son apparition aussi dans les années 1990 : l'ODS (Operational Data Store) est une base de données conçue pour centraliser les données issues de sources hétérogènes afin de faciliter les opérations d'analyse et de reporting. L'intégration de ces données implique souvent une purge des informations redondantes. Un ODS est généralement destiné à contenir des données de niveau fini par exemple un prix ou le montant d'une vente, en opposition aux données agrégées telles que le montant total des ventes. Les données agrégées sont stockées dans un DataWarehouse.

EIS : Executive Information System, terme apparemment utilisé dans les années 1990 pour désigner des outils de restitution d'information synthétique, souvent sous forme graphique, généralement seulement pour le top management. Contrairement aux outils Business Intelligence actuels, les EIS

fournissaient des rapports statiques. Il était quasiment nécessaire de faire un projet informatique dans les règles pour chaque rapport, c'est-à-dire faire un cahier des charges, un de spécification, le développement et le test. Il y avait donc un décalage important entre le moment du besoin et celui de la livraison.

D) Real-Time Data Warehousing (entrepôt de données temps réel)

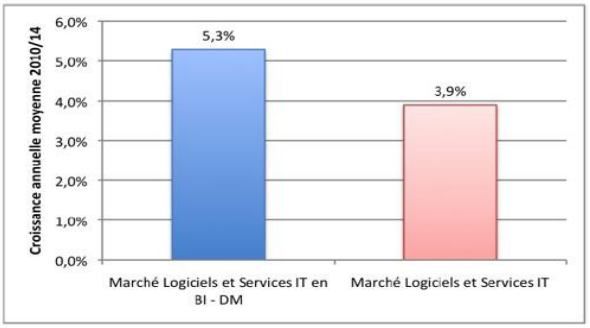
Notre prochaine étape (essentielle pour notre sujet) dans la saga de la Business Intelligence concerne le temps réel. Il faut maintenant éliminer de notre mentalité ETL qui a dominé depuis le commencement de cette histoire. La majorité des fonds investis dans la recherche et le développement était pour la récupération des données depuis les bases de données. Comment avoir un DataWarehouse qui lit les mêmes flux de données que les différents modules ? Comment ne plus avoir de data en file d'attente pour l'écriture dans le DataWarehouse ? Comment rendre le processus d'écriture indépendant du système d'exploitation ?
Voilà les questions qu'il faut se poser pour avoir un DataWarehouse temps réel. Mais il y en a encore d'autres.
Des solutions open source existent maintenant pour répondre à cette problématique. Teradata, leader historique, se voit en concurrence avec Bull (Europe), Datallegro (open source), etc.

E) Le marché

Il faut savoir que maintenant 95 % des entreprises sont équipées en décisionnel. Les logiciels et services liés à la mise en place d'outils d'intégration, de gestion, de manipulation, de restitution et de diffusion de l'information en Europe dépasseront le seuil des 2 milliards d'euros en 2011. Le marché devrait connaître une croissance moyenne annuelle supérieure à 5 % d'ici 2014. Le marché est en pleine évolution grâce notamment aux évolutions technologiques permettent un stockage et une manipulation importante de données. Le plus difficile n'étant plus de trouver l'information, mais de la traiter. L'information étant délivrée en temps réel, sous une forme adaptée aux nouveaux moyens d'accès et aux nouvelles habitudes de recherche sur internet. Elle arrive sur le secteur du décisionnel, notamment grâce à l'émergence du WEB 2.0 et des réseaux sociaux qui constitue un flux d'information exponentiel à analyser. On retrouve aussi les outils collaboratifs qui constituent aussi une part importante dans l'acquisition des données.

	Revenus 2008	Parts de marché 2008	Croissance 2007-2008
Application analytique et de gestion de la performance	3,055	34,7 %	24,3 %
Plates-formes BI	5,746	65,3 %	20,4 %
Ensemble	8,802	100 %	21,7 %

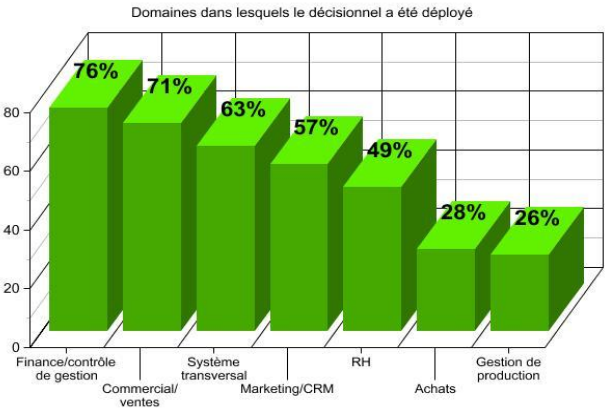
Revenus du marché du décisionnel dans le monde en 2008 (milliards de dollars)



Croissance annuelle du marché BI – DM vs. marché total
(Copyright PAC 2011)

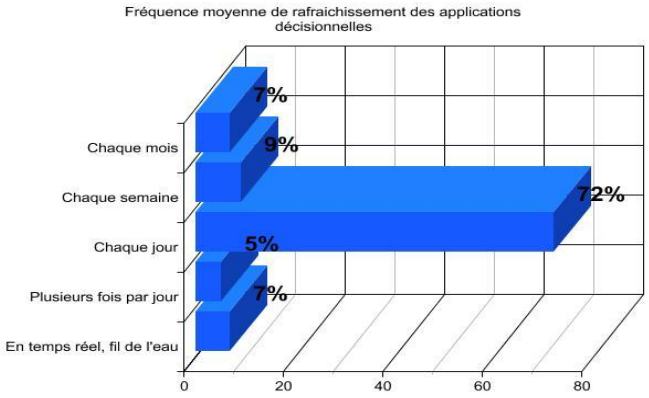
Croissance durant l'année 2010

Sur le marché du décisionnel, les acteurs sont présents dans tous les domaines d'activité, tous étant les grands de l'informatique comme Oracle, Microsoft, Amazon, SAP, etc.



Les applications décisionnelles sont majoritairement (76 %) mises en place pour répondre aux besoins dans les domaines de la finance et du contrôle de gestion. 33 % des répondants estiment d'ailleurs que s'ils ne devaient déployer qu'une application décisionnelle, ils privilégieraient les fonctions finance et contrôle de gestion. Le taux d'équipement en outils décisionnels est aussi fort dans le domaine commercial (71 %).

La possibilité de réaliser des mises à jour en temps réel vantée par les fournisseurs de solutions ne prend pas auprès des utilisateurs. Cette tendance demeure anecdotique (stable par rapport à 2008), et ne concerne que 7 % des répondants. Alors que les mises à jour quotidiennes restent le pilier essentiel avec 72 % des applications.



III. Acteurs de la Business Intelligence

Les DataWarehouses, ou plutôt en français, les entrepôts de données, sont devenus aujourd'hui un outil indispensable à la bonne marche d'une organisation. En effet, 95% du top 500 des entreprises américaines ont mis en place des entrepôts de données. A la base, le DataWarehouse était essentiellement destiné au marketing, mais il est maintenant un puissant outil de reporting aux entreprises.

A) Intérêt des DataWarehouses aux entreprises

La vocation du DataWarehouse est de donner aux entreprises une nouvelle façon de prendre des décisions autour de leurs activités majeures. Le principe du reporting prend tout son sens lorsque l'entreprise peut poser des questions (les « requêtes ») au DataWarehouse et obtenir des réponses. Ces réponses peuvent constituer des données qui permettront d'élaborer des documents, modèles graphiques, afin de « mesurer » le futur proche, ou de déterminer des associations entre deux produits, par exemple.
Il est évident qu'à partir de là, si on peut « prédire » le futur ou connaître les besoins des clients, à partir des données passées, qui sont donc stockées dans les entrepôts de données, les entreprises pourront prendre les meilleures décisions, et ainsi accroître leurs bénéfices ou encore donner une meilleure structure à l'entreprise pour accroître la productivité. Mais pas seulement, les interrogations aux DataWarehouses peuvent apporter d'autres éléments, pouvant être appliqués à beaucoup de domaines, et donc créer une *valeur ajoutée* à l'entreprise.
Bien que les entreprises appartiennent à de nombreux domaines différents, toutes ont un objectif, faire encore et toujours mieux. Le Business Intelligence le permet. Nous allons voir des exemples de sociétés.

B) Exemples de sociétés utilisant les DataWarehouses et la BI

1) Les chaînes de distributions

Wal-Mart, une grosse entreprise américaine de distribution grand public (équivalent à Carrefour en France) a prouvé les avantages du reporting par l'informatique en démarrant un projet de BI très peu de temps après la création de l'entreprise. Aujourd'hui, 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.

2) Les hôpitaux

En enregistrant leurs données dans le DataWarehouse, les hôpitaux disposent alors d'un système d'indicateurs de la santé de la population et de la qualité de soins. Les hôpitaux en Valais (Suisse) qui disposent d'un tel système, peuvent alors réaliser des études sur la santé publique (études épidémiologiques, prises en charge financiers, …).

3) En aviation

Les compagnies aériennes qui disposent d'un outil de reporting, peuvent identifier les agences de voyages qui contribuent fortement à leur chiffre d'affaires et donc peuvent décider de la mise en place d'un partenariat avec eux pour la plupart des destinations.
Ils peuvent aussi connaître les préférences de leurs clients, trouver des associations sur la plupart des destinations, réaliser un suivi sur le nombre de passagers par vol et déterminer – par exemple – si la fréquence de telle ou telle ligne peut être diminuée ou augmentée. En clair, l'idée est d'identifier tout ce qui peut être économisé, tout ce qui peut être bénéfique.
EasyJet a fait appel à la société Sopra (spécialiste en Business Intelligence) pour constituer un DataWarehouse. L'interrogation à cet entrepôt de données a permis de découvrir que la plupart des retards étaient imputables aux équipages, selon les rôles, les bases d'attache, les types de contrats !
Autrefois, il était impossible de comparer les performances d'un équipage basée à Rome et un autre à Londres, mais depuis que les informations sont cohérentes, les anomalies sont identifiables, les problèmes qui en résultent peuvent être résolus.

4) Energie

Sopra a été retenu par EDF pour la mise en place « Business Intelligence » qui consistera à faciliter la consolidation et la disponibilité de l'information, découvrir et exploiter les référentiels communs partagés entre les différents acteurs du nucléaire (ou de l'électricité), et favoriser les meilleurs pratiques du métier. Il s'agit ici d'établir un « tableau de bord » performant et disponible rapidement (Temps réel).
Une autre application, plus générale concerne l'optimisation de la consommation d'énergie. En collectant les données sur la consommation énergétique, on peut converger sur les causes de consommation superflues (pour ne pas dire « surconsommation »), qui peuvent être alors réduites.

Partie 2 : Axes de recherche

I. Cloud BI

A) Le principe du Cloud et la Cloud BI

Une solution de Business Intelligence basée sur le Cloud est une des nouvelles tendances du marché de l'informatique. Elle s'appuie sur la généralisation et le développement du Cloud Computing.

1) Le Cloud Computing

Ce terme désigne la virtualisation et la mutualisation des ressources et des services. Les données et les services sont déportés sur des parcs de serveurs distants et non plus stockés ou exécutés en local ou sur le poste utilisateur. Selon le National Institute of Standards and Technology *(NIST)*, le Cloud Computing est l'accès via le réseau, à la demande et en libre-service à des ressources informatiques virtualisées et mutualisées. Il faut différencier les trois types de Cloud suivants :

- Le Cloud privé interne : il s'agit d'un Cloud géré et hébergé en interne par l'entreprise propriétaire.

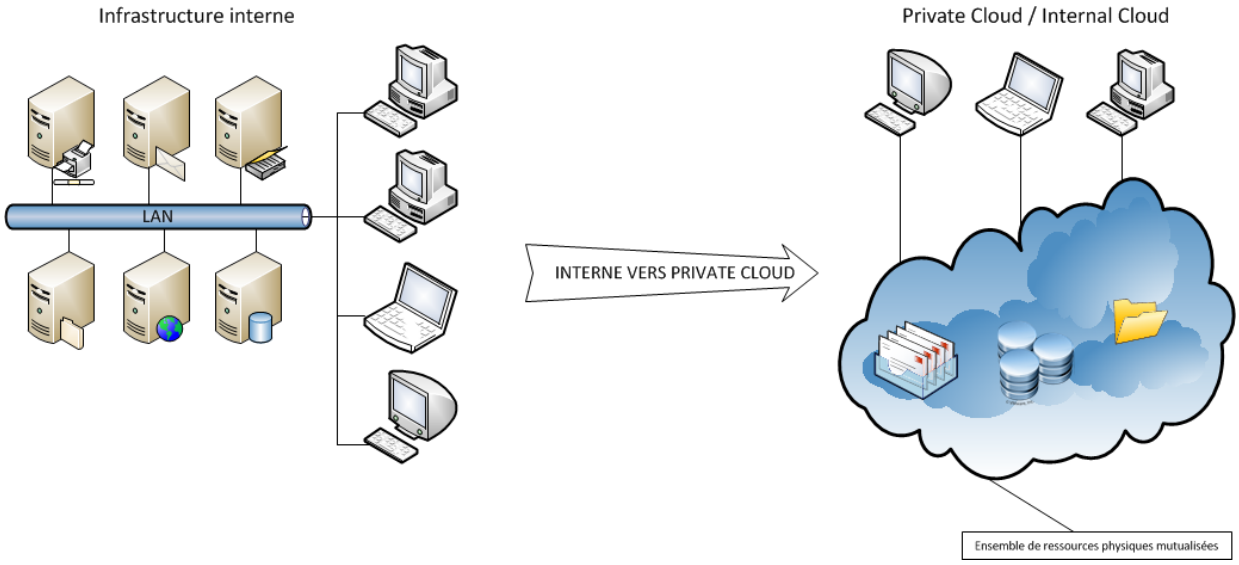


Figure 1 - Le passage d'une architecture classique au Cloud privé

- Le Cloud privé externe : un prestataire fourni à un l'entreprise un Cloud qu'elle seule utilisera mais la gestion échoie au fournisseur et non au client.

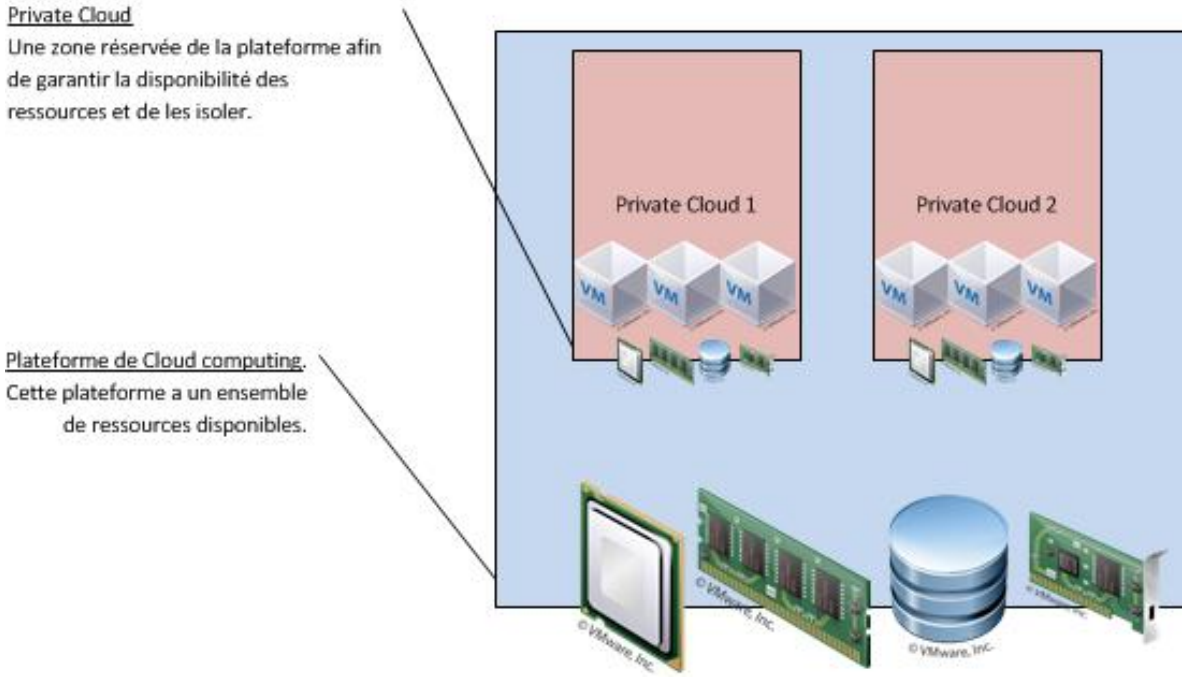


Figure 2 - Le Cloud privé externe

- Le Cloud publique : ensemble de ressources géré par des entreprises spécialisées qui proposent leur service à d'autres, qu'elles soient utilisatrices du Cloud ou elles-mêmes prestataires de service sur ce Cloud.

Enfin le Cloud hybride n'est pas à proprement parlé un type de Cloud mais plutôt une organisation spécifique permettant de mettre en communication des Clouds publiques et privés. Le défi de cette organisation est la communication entre des types d'infrastructures différentes.

Le Cloud Computing offre une bonne fiabilité de service compte tenu de ses infrastructures dotées d'une bonne tolérance aux pannes (le plus souvent grâce à la présence de répliques sur le Cloud). Bien que la fluidité d'accès au service soit tributaire de la bande passante, ce n'est pas un problème du fait de l'internet haut débit.

Une fois les données injectées dans le Cloud il est impossible de savoir précisément où elles se trouvent. Elles peuvent très bien être réparties sur plusieurs serveurs du parc et peuvent également être relocalisées selon les besoins et politiques de gestion du Cloud.

2) La Cloud BI

Le terme Cloud BI est souvent utilisé pour rassembler l'informatique décisionnelle en mode SaaS *(Software as a Service)*, en mode PaaS *(Platform as a Service)* et les applications SaaS offrant des fonctionnalités spécifiques de Business Intelligence. De plus, l'informatique décisionnelle peut également être présente sur le Cloud en tant que IaaS *(Infrastructure as a Service)*. Il est donc important de spécifier ces termes différemment :

- SaaS : le client accède à un logiciel d'application auprès d'un fournisseur de solution. Ce service est présent sur les Clouds publiques.
- PaaS : le fournisseur de la solution BI développe, maintient, exploite et héberge la solution pour le client (Clouds publiques ou privés externes où la solution BI est gérée par un prestataire spécialisé).
- IaaS : le client loue l'accès au matériel en déployant son système de BI sur des serveurs distants. Seule une partie limitée de la gestion du serveur échoue au client (Clouds privés externes).

La Cloud BI permet à une entreprise de mettre en place son système de Business Intelligence sans avoir à installer et implémenter de logiciel au sein de son système d'information. L'entreprise n'a plus besoin d'investir dans du matériel ou des licences de logiciel (excepté pour le Cloud privé interne). Les serveurs et applications sont à la charge du prestataire mettant en place le Cloud. Ce type de BI est donc moins couteux que ses homologues (qui requièrent un investissement initial souvent important). Le prestataire, quant à lui, exploite son parc de ressources en offrant à plusieurs entreprises d'exploiter une même ressource, chacune pour ses propres besoins.

Les bases de données relationnelles, transactionnelles, les autres types de données et les logiciels de BI constituant le système décisionnel de l'entreprise n'utilisent donc pas les ressources de celle-ci (que ce soit en termes de stockage ou de puissance de calcul) mais celle du Cloud. L'investissement de l'utilisateur de la Cloud BI ne concerne en conséquent que ce qu'il utilise (nul besoin d'investir dans davantage de matériel afin de prévenir des éventuelles augmentations des charges de données ou de traitements). C'est pourquoi les prestataires de solutions Cloud BI proposent des forfaits permettant à un certain nombre d'utilisateurs d'accéder à leur système décisionnel en ligne via leurs navigateurs web. Le client paie pour ce qu'il utilise uniquement (que ce soit le nombre d'utilisateurs, les ressources nécessaires où le type de service auquel il souscrit). De plus, en cas d'accroissement des besoins du système supportant les outils et sources du système décisionnel, le Cloud sur lequel ceux-ci fonctionnent réassigne automatiquement des ressources (flexibilité et élasticité).

Les utilisateurs n'ont pas besoin d'une connaissance étendu au niveau technique et informatique. Les outils de reporting sont faciles d'accès et d'utilisation. Ils peuvent également être personnalisés. Ce type de système décisionnel ne requière donc plus de larges connaissances de la part de ses utilisateurs ce qui permet au client de diminuer son investissement pour la formation de ses employés. Tous les utilisateurs possédant les identifiants requis par les applications du système BI s'exécutant sur le Cloud peuvent y accéder et, selon le prestataire, échanger les rapports BI qu'ils ont extraient de leurs données.

Enfin, l'un des avantages de la Cloud BI est l'accès à de grandes bases de données publiques elles-mêmes présentes sur les Clouds publiques. Ces données peuvent être associées à celles de l'entreprise ce qui permets d'extraire des tendances que seules les données de l'entreprise n'auraient pu fournir. Les grands noms hébergeant ces bases de données sont Google analytics, Salesforce ou encore Amazon web service. Ces mêmes hébergeurs proposent aux entreprises et institutions de rendre la consultation de leurs bases de données stockées sur le Cloud à tous, cependant cela n'est pas obligatoire.

B) Quels utilisateurs et pourquoi ce choix ?

Le domaine du Cloud BI a beaucoup évolué et propose aujourd'hui des interfaces simples et facilement accessibles. La majorité des entreprises choisissant ce type de BI sont des PME bien que quelques grandes entreprises aient opté pour cette solution. Le choix du Cloud BI est en général lié à un budget BI faible ou diminué, une grande fragmentation territoriale des agences constituant l'entreprise ou une volonté de moins dépendre des services informatiques internes à l'entreprise.

L'investissement de base dans la mise en place d'un système BI n'étant pas abordables pour toutes les entreprises, la Cloud BI offre une alternative aux Petites et Moyennes Entreprises. Le coût total de possession ne concerne que les postes de travail permettant d'accéder aux services du prestataire (un ordinateur ayant un accès internet est suffisant). La Cloud BI ne génère donc pas de cout d'investissement mais des couts de fonctionnement (forfait prestataire). En outre la maintenance n'est plus à la charge de l'entreprise.

Un autre avantage de la Cloud BI est son accessibilité depuis quelque lieu que ce soit (pourvu qu'un accès internet soit disponible). Ainsi, une entreprise ayant ses différents services et filiales géographiquement éclatés peut permettre à ses employés de travailler sur les mêmes données présentes sur le Cloud. Le haut débit permet un temps d'attente relativement faible, cependant plus le centre hébergeur du Cloud est physiquement éloigné, plus le temps de réponse est important. Le vendeur de pièce automobile américain AutoZone a ainsi choisi une solution Cloud BI pour ses 4000 points de ventes. Les données générées par cette multitude de source est donc exploitables au niveau local pour les employés des différents points de vente, et combinées elles sont utiles à la direction de l'entreprise.

L'un des freins majeurs reste la sécurité ou la satisfaction des entreprises par rapport à leurs systèmes actuels de BI. En effet la sécurité des données est un point important puisque les entreprises utilisent leur système décisionnel pour rester compétitives. La BI appuyant la prise de décision stratégique, les données confidentielles doivent donc être sécurisées. Les outils de Cloud BI sont accessibles quelque soit l'endroit pourvu que l'utilisateur ait en sa possession des identifiants lui permettant de s'authentifier auprès du service prestataire, et les données de l'entreprise sont stockées sur le Cloud (sans qu'elle sache précisément où). Cela pose le problème des différences de législation entre les pays où sont présents les différents serveurs constituant le Cloud. En effet le client ne sait pas où sont entreposées ses données et dans le cas de la législation chinoise, le gouvernement chinois peut demander à consulter les données présentes sur un serveur situé sur leur territoire sans avoir à en informer l'entreprise propriétaire (le client du service de Cloud BI). De plus

les clients de la Cloud BI restent inquiets d'une éventuelle disparition de leurs données (certains accidents ayant entrainé des pertes laissent les entreprises sur leur garde). Enfin les éventuels clients de la Cloud BI questionnent l'avenir de leurs bases de données dans le cas ou ils mettraient un terme à leur contrat avec un prestataire de Cloud BI (la base de données et ses répliques seront-elles bien effacées).

L'autre frein à la mise en place de la Cloud BI est la satisfaction des clients par rapport à leurs systèmes actuels. Les utilisateurs de BI classiques ont investis dans leur système décisionnel et devraient alors laisser derrière eux une grande partie de leur investissement alors que leur système les satisfait. En outre un changement pour une organisation Cloud BI instaure une grande dépendance à internet puisque sans réseau les outils sont inaccessibles.

C) Les acteurs majeurs et les nouveaux intervenants

Les trois leaders de la Business Intelligence, SAP, Oracle et SAS s'intéressent également aux nouveaux marchés qu'offre la Cloud BI. Chacune des trois entreprises propose sa solution de Cloud BI à la demande.

SAP BI OnDemand

SAP propose une large variété de logiciels de BI avec la suite intégrée *Business ByDesign*, des logiciels métiers tels que *Sales OnDemand* ou encore *Expenses OnDemand* et des logiciels de BI comme *BusinessObjects BI OnDemand*. Avec ces différents logiciels SAP vise à la fois les grandes entreprises et les PME. L'intégration des données et le stockage de celles-ci dans des entrepôts de données est supportée dans le Cloud. Des outils de reporting et de monitoring sont disponibles en ligne. Le client n'est responsable que de la conception du DataWarehouse, des Univers, des scripts de Data Integrator, du contenu BI et de l'administration des comptes utilisateurs et des permissions sur les documents. SAP est associé au Cloud Amazon mais propose également une application CRM (gestion des relations client) en partenariat avec Salesforce. La solution RH sur le Cloud de SAP (*SAP Core HR*) sera lancée en mai 2012.

SAP OnDemand se présente sous trois packages de souscriptions différents dont les tarifs vont d'une totale gratuité à 63 euros par mois et par utilisateurs.

Oracle : un éventail de solutions

Oracle propose différentes solutions de Cloud BI :

- Des SaaS, middleware et bases de données sur l'*Oracle public Cloud*.
- La gestion de services en mode Cloud (déploiement, gestion, sécurité, …) qu'ils soient hébergés par Oracle ou par le client.
- La mise en place d'un Cloud privé basé sur les dernières technologies Oracle (applications, plateforme de développement et infrastructure Cloud d'Oracle). Cette solution permet au service informatique de rester maitre du système décisionnel (sécurité, conformité, qualité de service et fonctionnalités disponibles).

La solution Oracle pour la gestion et BI des ressources humaine : *Oracle Fusion HCM Cloud Service* est disponible sur le Cloud publique d'Oracle.

SAS OnDemand Business Intelligence

SAS OnDemand propose différentes applications de type SaaS ainsi que des solutions d'hébergement. Ces applications sont associées à des services d'experts métier et basé sur la plateforme SAS.

La solution Cloud de SAS pour la Business Intelligence permet aux utilisateurs d'accéder à des outils de reporting et de gestion en ligne. Cette solution de BI se base sur des cubes OLAP et des bases de données relationnelles. En outre SAS OnDemand BI intègre Microsoft Office.

IBM et Cognos

IBM SmartCloud Application Services est une nouvelle PaaS dont une des application est dédiée à la gestion des environnements SAP. Avec cette nouvelle plateforme sur le Cloud IBM permet une diminution drastique des temps d'exploitation d'une solution SAP (le temps clonage de base de données passe de 2-3 jours à une vingtaine de minutes par exemple).

En outre il est possible de mettre en place Cognos sur un Cloud tel que l'Amazon Elastic Compute Cloud. Il ne s'agit pas d'un service de la part d'IBM mais de la mise en place d'un système décisionnel sur un Cloud publique, il est donc nécessaire de posséder une licence Cognos.

Bien que les acteurs majeurs de la BI traditionnelle se soient lancés dans les solutions de Cloud BI, de nombreuses autres sociétés se sont également créées suite au développement du Cloud Computing.

Jaspersoft (US)

Cette société propose une plateforme décisionnelle et travaille en partenariat avec des éditeurs de SaaS et d'intégration des données afin de proposer des outils de reporting performants, simples d'utilisation et personnalisables. En outre, Jaspersoft propose une intégration SAP. Jaspersoft compte en 2012 plus de 100 clients pour sa solution de Cloud BI.

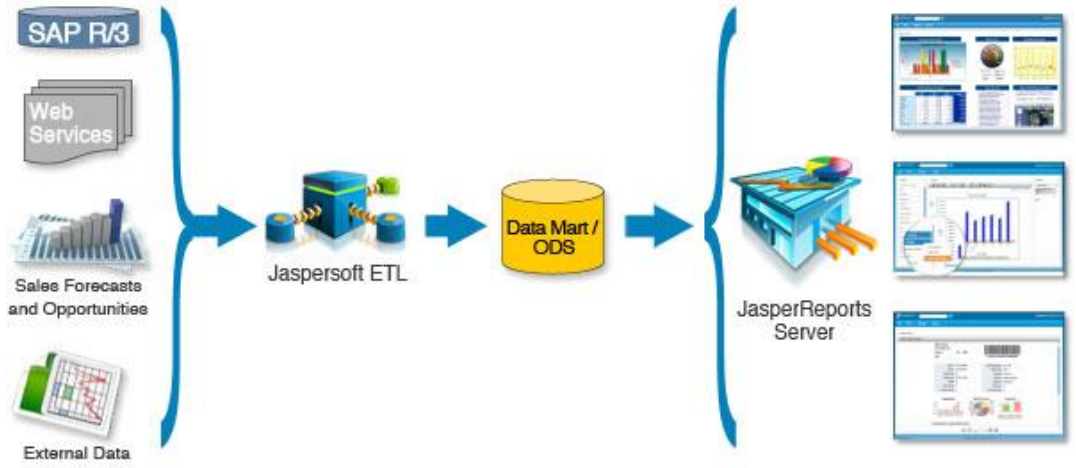


Figure 3 - Jaspersoft et son intégration SAP

- ***GoodData (US)***

Ce prestataire offre une solution flexible proposant le Cloud BI comme un service internet de type SaaS. Les tableaux de bords sont personnalisables selon les besoins de l'entreprise cliente et les variables analysées peuvent également être personnalisées. La solution de BI s'accompagne d'un outil collaboratif afin que les différents acteurs des décisions puissent partager et discuter les résultats des rapports. Cette solution propose aussi des applications très compétitives permettant l'accès à des sources de données publiques telles que Google Analytics ou Zendesk. GoodData est utilisée par de grandes sociétés telles que PSA ou CapGemini.

- ***BIRT onDemand (US)***

La solution de BIRT est un SaaS offrant des tableaux de bord simple et ergonomique. Une fois les bases de données créées ou exportées vers les serveurs RDS (remote desktop service), les outils de BI et de reporting sont accessibles via un navigateur web. Les prix de cette solution varient de 22 euros à 30 euros par mois.

- ***BIME (FR)***

Cette solution exploite le cloud Amazon et propose de nombreux tableaux de bords et outils de reporting. Il s'agit également d'un SaaS. La solution BIME permet la connexion à une base de données pré existante et son exploitation dans un navigateur web. Cette solution a été adoptée par Arcellor Mittal et par des agences de consulting et d'e-commerce. La version 3 de BIME offre un QueryBlender permettant de lier des données provenant de différentes sources (en ligne ou hébergées en local). Cette nouvelle version s'approche d'un système de base de données fédérées et permet également de consolider différentes sources d'informations (privées ou publiques telles que Google Analytics, Salesforce ou encore Amazon web services). Les prix varient de 45 euros à 180 euros par mois par analyste.

- ***BITTLE (FR)***

Ce prestataire propose un SaaS s'appuyant sur Google App Engine. L'analyse peut porter sur des fichiers excel, csv ou xml mais également sur des bases de données existantes. Les outils de reporting sont accessibles via un navigateur web. Les prix varient de 0 euros à 149 euros par mois et par utilisateur.

Les éditeurs proposent tous différents systèmes de Cloud BI. Le choix de l'entreprise doit donc prendre en compte son budget, ses besoins en BI et les données (tout ou partie) qu'elles souhaitent transférer sur le Cloud. La Cloud BI, associée à la mobile BI (cf. Partie 2 – VIII) peut donner à une entreprise un avantage de réactivité pour ses prises de décision et une plus grande marge de manœuvre pour sa saisie et son monitoring des données.

II. Compression des données

A) Compression des données

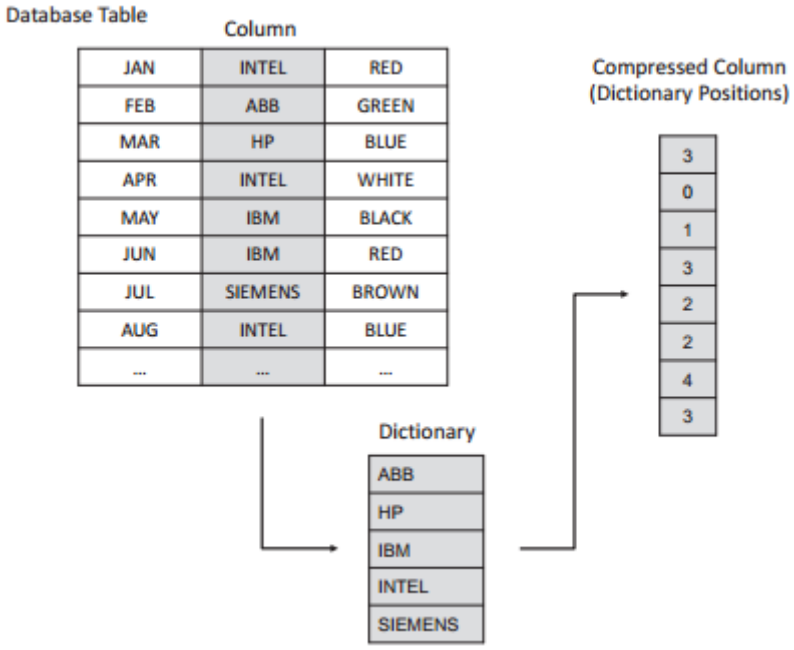
La compression des données est une partie essentielle pour le BI temps réels. Que ce soit pour la rapidité d'exécution ou la consommation mémoire. Elle est maintenant au cœur de toutes les grandes bases de données. L'augmentation de la capacité mémoire des nouveaux calculateurs permet à l'heure actuelle un calcul pour rapide. Cependant, son coût reste jusque-là conséquent pour traiter un volume important de donné (nous parlons ici de plusieurs téra byte de données), le recours à la compression en mémoire est alors essentiel dans de tels systèmes.

Les bases de données orientées colonnes sont particulièrement visé pour la compression de données, en effet celle ici contienne des sections consécutives de données de même types. Il est alors possible d'utiliser des algorithmes de compression, spécialement conçue pour ce type de base de données.

Le couple puissance du CPU et compression augmente considérablement le temps d'exécution des requêtes. Il est important de préciser qu'une augmentation de la bande passante CPU <-> mémoire pour une compression dite "légère" n'apporte pas de gain significatif. A contrario d'une compression dite "lourde" où l'accès mémoire est plus important. Dans cette partie nous décrirons les différences qu'il peu y avoir entre une compression "légère" et une compression "lourde" ainsi que leur impacte sur une base de données en colonnes. Nous terminerons cette section par divers exemples.

1) Compression dite "légère" : Light-weight

La plus pars des méthodes de compression légère sont basées sur un système de dictionnaire. En pratique chaque valeur distincte d'une colonne est ajoutée au dictionnaire. Ainsi chaque occurrence dans la base de données correspondant à cette valeur est remplacée par la position de cette occurrence dans le dictionnaire. Un exemple sera beaucoup plus parlant.

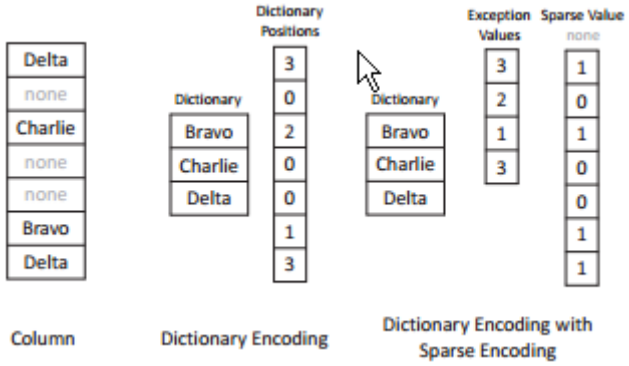


Compression à l'aide d'un dictionnaire

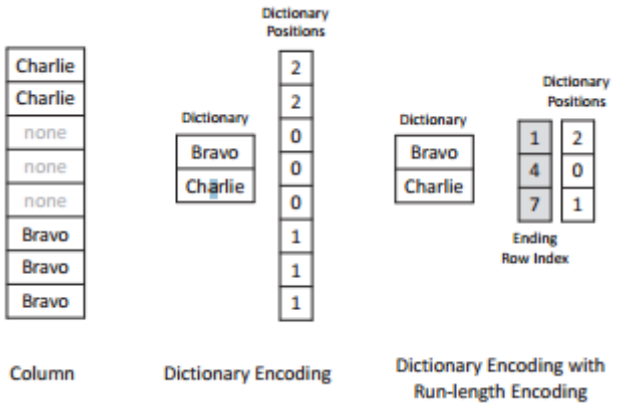
Dans la plus pars des scénarios réels ce type de compression est très utile. En effet, il est beaucoup plus rapide de chercher si la valeur existe dans le dictionnaire que de recherche dans la base de données entière. Nous détaillerons brièvement par la suite deux techniques de compression de données à l'aide de dictionnaires. Il est existe d'autre, mais celle-là nous semble essentiel à savoir "Sparse encoding", "Run-Length Encoding".

- Sparse Encoding

Sparse Encoding est une technique de compression par dictionnaires (contenant des données de valeur null et non null) qui va s'appuyer sur une colonne supplémentaire appelée "Sparse Value". Cette colonne reprend les mêmes index de ligne que la colonne que nous cherchons à compresser. Il est alors possible de définir si la valeur stockée est contenue dans le dictionnaire à l'aide d'un simple booléen. Si cette donnée existe dans le dictionnaire nous alors recherche ça valeur dans la table des exceptions. Ici nous travaillons sur des types de données beaucoup plus courtes (boolean, integer) que les données stockées dans la base.



Sparse Encoding



Run-Length Encoding

- Run-Length Encoding

Run-Length Encoding exploite le fait qu'une valeur peut être la même d'une ligne à l'autre (consécutivement). Cette technique de compression à l'aide de dictionnaire compte le nombre de répétitions successives d'un terme. La position de ce terme est alors placée dans une table de position. Puis nous utilisons une table de fin de position pour déterminer quand cette valeur (qui est consécutive à plusieurs lignes) se termine. Dans l'exemple si dessous nous voyons que "Charlie" possède deux occurrences de suite, le terme "nul" trois occurrences de suite et le terme "Bravo" trois occurrences de suite. Nous déterminons alors conjointement là table de positions et la table de fin de position. "Charlie" se termine en second ligne donc position 1 dans la table (le 0 étant la première position). "Null" se termine à la ligne 5 donc positions 4 dans la table de position, etc.

2) Différence de performance

Voici un exemple de performance que nous retrouvons sur une base de données client de type ERP. Utilisant des dates, des numéros de carte de crédit, le prix et une valeur en %. Cette comparaison se fait entre la méthode Sparse et la méthode Run-Length. Cet exemple nous permet de voir qu'il est important d'adapter son modèle de compression suivant le type de données contenu dans notre base.

	Sparse	RLE	Indirect
Amount	2.5	2.7	**4.9**
Date	222.3	**1948.1**	231.1
Discount	**113**	103.1	106.2

Performance de compression suivant différentes compressions de type légères

Nous nous rendons compte par exemple que seule la compression à l'aide du modèle indirect nous permet d'avoir un gain significatif pour les valeurs réelles et courtes (exemple des prix : Amount). On option une colonne 80% plus petite seulement avec le système d'encodage de type Indirect. Autre exemple pour l'utilisation d'une base de données contenant des valeurs null (ici le cas de la table discount) la méthode Sparrse reste la meilleure. Alors qu'une colonne contenant des Dates diminue très largement en poids comparer à toute les autres méthodes de compression légère de l'ordre de 180%.

3) Compression dite "lourde" : Heavy-weight

L'algorithme le plus populaire de compression de données lourde est appelé : Lempel-Zip. L'idée ici est de remplacer les occurrences identiques par là référence de l'occurrence précédente. Cette table de référence est créée automatiquement pendant que nous travaillons sur les données. Le problème de ce type de compression est qu'il demande plus de puissance de calcul qu'un algorithme de compression légère. Il existe aussi des algorithmes de compression dite hybride qui utilise les deux approches.

4) Partitionnement horizontal

Le partitionnement horizontal divise une table en plusieurs tables. Chaque table contient alors le même nombre de colonnes, mais moins de lignes. Par exemple, une table qui contient 1 milliard de lignes peut être segmentée horizontalement en 12 tables, chacune de ces tables représentant un mois de données d'une année donnée. Toute requête recherchant les données d'un mois spécifique ne consulte que la table concernée.
Le choix du mode de partitionnement horizontal des tables dépend de la façon dont les données sont analysées. Il convient de segmenter les tables de sorte que les requêtes consultent le moins de tables

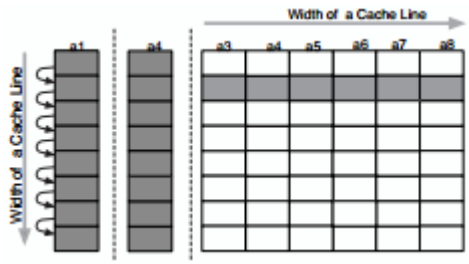
possible. Sinon, le recours à un trop grand nombre de requêtes UNION pour fusionner logiquement les tables lors de la requête risquerait d'affecter les performances.
Le partitionnement horizontal des données basé sur l'âge et l'utilisation est le plus fréquemment utilisé. Une table peut contenir, par exemple, des données concernant les cinq dernières années, mais seules les données concernant l'année en cours sont régulièrement consultées. Dans ce cas, vous pouvez envisager le partitionnement des données en cinq tables, chacune contenant les données relatives à une seule année.

5) Partitionnement vertical

Le partitionnement vertical divise une table en plusieurs tables contenant moins de colonnes. Les deux types de partitionnement vertical sont la normalisation et le fractionnement des lignes.

La normalisation désigne le processus de base de données standard consistant à supprimer les colonnes redondantes d'une table et à les replacer dans des tables secondaires liées à la table primaire par des relations de clé primaire et de clé étrangère.

Le fractionnement de lignes divise verticalement la table initiale en tables comportant moins de colonnes. Chaque ligne logique dans une table fractionnée correspond à la même ligne logique dans les autres tables et identifiée par une colonne UNIQUE KEY identique dans toutes les tables partitionnées. Par exemple, la jointure de la ligne avec ID 712 de chaque table partitionnée reconstitue la ligne initiale.



Exemple de partitionnement vertical

Comme le partitionnement horizontal, le partitionnement vertical permet d'analyser moins de données lors des requêtes. Les performances de requête s'en trouvent ainsi améliorées. Par exemple, une table comportant sept colonnes, dont seules les quatre premières font l'objet de consultations régulières, bénéficiera du fractionnement de ses trois dernières colonnes dans une table séparée.

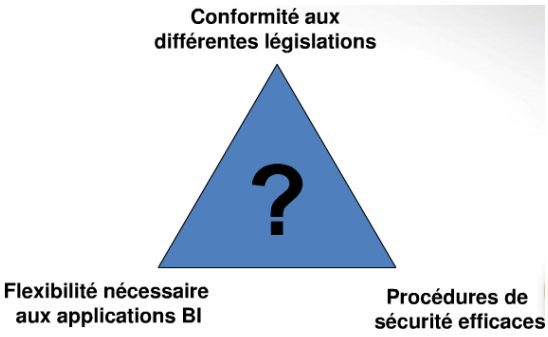
Le partitionnement vertical doit être envisagé avec précaution, car l'analyse des données provenant de plusieurs partitions nécessite que les requêtes joignent les tables. Les performances peuvent également pâtir du partitionnement vertical si les partitions sont très volumineuses.

III. Techniques de sécurisation des données

Pourquoi sécuriser la BI ?

L'enjeu de la sécurisation de son système informatique est pour l'heure au centre de toutes les problématiques d'entreprise. Sur un système BI, on peut se demander pourquoi sécuriser les données. Plusieurs réponses s'offrent à nous tout d'abord le risque d'affaires (externe) pour ne pas divulguer les objectifs à la concurrence. En second nous retrouvons les risques organisationnels (interne) qui consistent à ne pas divulguer des informations à des salariés qui ne doivent pas y avoir accès. On retrouve un autre risque majeur celui du risque réputationnel, par exemple : divulgation des données clients au monde extérieur qui nuirait à l'image de l'entreprise. L'enjeu de la sécurisation répond à un dernier risque, le risque opérationnel par exemple la corruption des données.

Plusieurs organismes mondiaux par exemple pour l'Europe : EU data protection laws. Mettent en place différentes lois obligeant les sociétés à un minium de sécurité sur leurs systèmes d'information. Comment adapter son système BI pour plus de sécurité en rapport à ces lois ? Comment réduire la perte de temps que pour engendrer ce genre de procédure sur notre système temps réel. Cet axe répondra à ces différentes questions.



Nous citerons en exemple : Le processus de reporting est dans 70% des cas un processus manuel non automatisé. Pour cela, l'utilisateur a accès à la base de données puisqu'il a besoin d'écrire des requêtes SQL spécifiques ce qui pose un premier problème en matière de sécurité. Nous avons affaire ici à un environnement collaboratif.

Pour la sécurité d'un système BI, il est important de mettre en place les 5 axes si dessous :

- Définition d'une politique de sécurité
- Procédures de sécurité efficaces
- Gestion de la sécurité logique
- Gestion de la sécurité physique
- Audit périodique du système

A) Définition

Une politique de sécurité est un plan d'actions définies pour préserver l'intégrité et la pérennité d'un groupe social. Dans la plus pars des cas elle est signée par le CEO, elle implique donc la haute direction. Elle contient la définition des exigences minimales de sécurité et de confidentialité pour la conception, l'implémentation, administration et l'utilisation de la BI au sein de l'organisation. Tous les membres de l'organisation doivent avoir pris connaissance de la politique et la signer. Elle implique la responsabilité sur le plan individuel.

B) Sécurité logique

La sécurité logique consiste à garantir l'accès aux seules personnes autorisées et uniquement à l'information qui leur est destiné. Pour cela, nous avons besoin de mettre en place :

- une authentification de l'utilisateur (mot de passe)
- un contrôle d'accès
- un chiffrage des données pour éviter toute attaque interne comme externe

Pour l'authentification il faut s'assurer que le serveur BI est compatible avec différent type d'annuaires d'entreprise (LDAP, Active Directory, Etc.)
Pour le contrôle d'accès nous pouvons mettre en place des procédures visant à l'imité d'accessibilité de l'information par l'utilisateur. Tous les différents systèmes BI que ce soit Oracle VPD, Business Objects (universe) ou Cognos possèdent dans leur configuration cette approche.
Quant aux chiffrages des données, c'est l'une des meilleures défenses que nous pouvons mettre en place sur un environnement BI. Surtout si celui-ci est accessible par le WEB. Par exemple elle devient retrouvable lors d'une attaque passive.

C) Sécurité physique

La sécurité physique concerne les serveurs (data centers) et les postes clients (utilisés pour administrer et gérer les applications BI). Pour palier à problème, il est important d'utiliser un script interne à l'entreprise qui va vérifier périodiquement la validité de la sécurité des environnements et des applications BI. Pour cela nous pouvons mettre en place un IDS (Intrusion Detection System).
À l'heure d'aujourd'hui les informations des environnements BI sont devenues disponibles à partir de multiple application client : laptop, téléphones cellulaires, etc. Ce genre d'applications augmente le risque de façon inégalée depuis ces dernières années. Il est donc important là aussi d'utiliser un chiffrage sur les applications client.

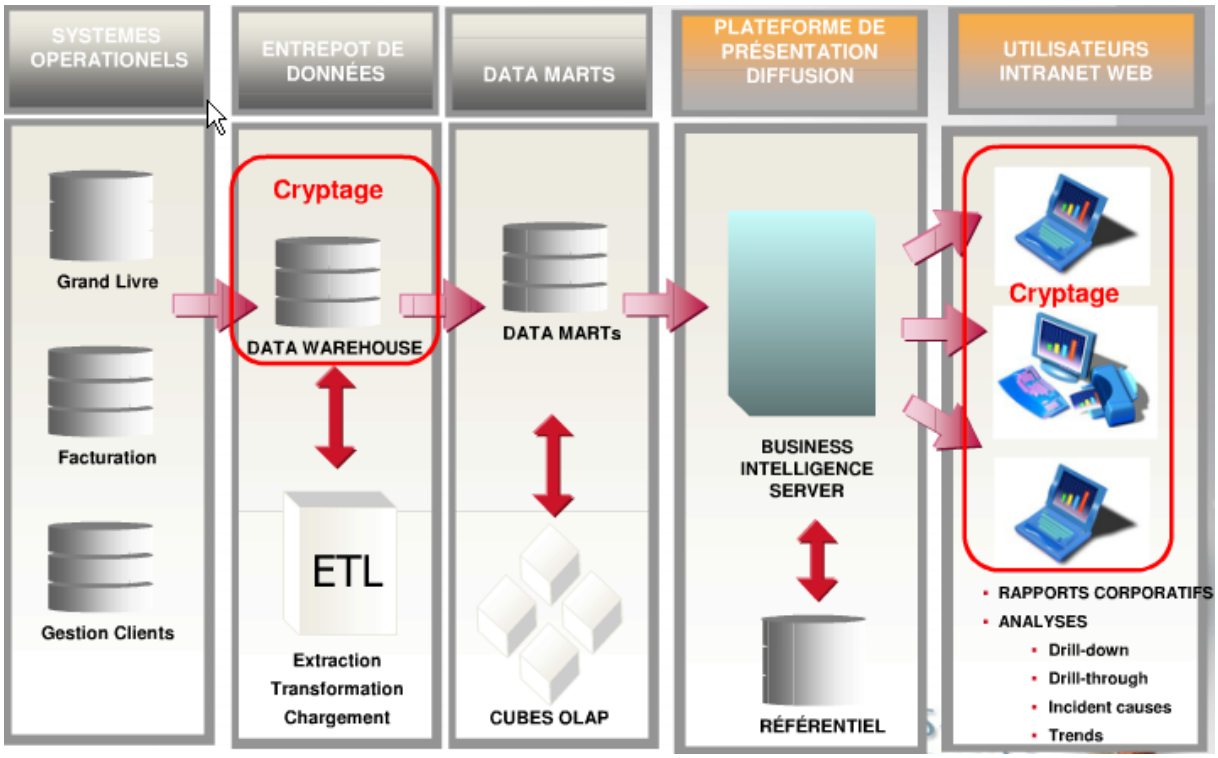


Schéma chiffrage des données sur un système BI

Il n'existe pas de solution ultime pour sécuriser un système BI toutes les solutions décrire repose sur des "Best pratice".

IV. Solutions matérielles

A) Solutions matérielles permettant de faire la business intelligence temps réel

Les entreprises qui souhaitent faire de la business intelligence en temps réel devront bien sûr s'acquérir un matériel le permettant, même si le Business Intelligence dépend largement du côté software et de l'architecture.

B) Quel matériel acquérir ?

Tout dépend des besoins de l'entreprise, ainsi que sa dimension. Par exemple, en ce qui concerne la capacité de stockage, l'entreprise va-t-elle enregistrer un nombre relativement peu élevé, ou au contraire extrêmement élevé ?

Wal-Mart, une chaîne de grande distribution, va atteindre dans quelques années le Pétaoctet, ce qui représente près de 1000 Téraoctets. Elle fait partie des entreprises qui stockent le plus de données au monde. Les besoins de cette grande distribution ne sont évidemment pas le même que pour une entreprise de dimension plus réduite.

Ensuite, le Business Intelligence en Temps Réel diffère du Business Intelligence « classique », il faut traiter et disposer le plus rapidement possible les informations : un matériel performant est donc nécessaire.

C) Quelles sont donc les matériels qui permettraient de faire du Business Intelligence Temps Réel ?

- ***Serveurs parallèles de Teradata***

Teradata, constructeur et éditeur des solutions informatiques, et spécialisé dans l'entreposage des données et les traitements analytiques, propose des serveurs massivement parallèles (« Massively Parallel Processing »), qui sont conçus selon une architecture sans partage (« Shared Nothing Architecture »), qui intègre des serveurs Intel, et une interconnexion à haute vitesse, le BYNET.

Les machines Teradata se basent souvent sur un système d'exploitation spécifique, le « NCR UNIX SVR4.2 MP-RAS », qui est une variante d'un système appartenant à AT&T (dont Teradata s'était séparé en 2007)

Toutefois, elle peut fonctionner sur des serveurs d'Intel 64-bits sous d'autres systèmes d'exploitations, comme Windows 2000, Windows Server 2003, SuSE (Linux).

Les entrepôts de données de Teradata sont alimentés en quasi-temps réel, principalement grâce à des applications spécifiques fondés sur l'ETL, auxquelles on accède via le ODBC, ou le JDBC.

Teradata a pour avantage d'être l'un des plus grands entreprises spécialisés, si ce n'est le leader, ce qui fait que la grande majorité des éditeurs du décisionnel proposent des solutions compatibles avec les systèmes Teradata.

Technologie d'exécution en mémoire (« in-memory ») de SAP

Pour disposer des rapports et tableaux de bord en temps réel, il y a les problématiques de performance, en effet, la complexité des bases multi-dimensionnelles et l'accumulation des données peuvent « ralentir » , voire ne plus donner au terme « temps réel » tout son sens.

Donc, pour permettre d'accélérer le temps de traitement, on peut utiliser les technologies d'analyse en mémoire (« in-memory ») qui consistent à stocker les données en mémoire vive. De même on peut tirer parti des architectures multicores, pour optimiser l'exécution des requêtes.

SAP propose cette technologie « SAP In-Memory Appliance » (SAP HANA). Notons que ceci est un logiciel, mais elle utilise la mémoire vive. D'où la possibilité d'acquérir des mémoires vives de capacités plus larges.

Le System z de IBM

Le système z désigne tous les ordinateurs centraux fabriqués par IBM.

Par exemple, le serveur z10 est l'équivalent à, d'après IBM, 1500 serveurs tournant sous un système x86, tout en consommant 85% de puissance électrique en moins, de même en volume. Le serveur utilise 64 processeurs simultanément. D'autres serveurs de cette série existent, comme le zEnterprise196, qui utilise 96 processeurs à 5,2 GHz, ce qui permet de supporter 100.000 images virtuelles.

Cognos, la solution logicielle d'IBM pour la Business Intelligence, peut être installée sur ces serveurs.

V. Techniques d'exploration et de restitution des données

En 2012, on ne peut plus parler d'exploration des données sans parler de « **Big Data** », qui est le nouveau défi en Business Intelligence : la question est de savoir quoi faire de l'énorme flot de données récupérés via

- Les bases de données internes.
- Le suivi des consommateurs sur
  - les sites : leurs achats, leurs comportements, leurs avis.
  - les réseaux sociaux : leurs habitudes et les liens qu'entretiennent les consommateurs entre eux.

En effet de plus en plus d'informations sont récupérées et ce à tout instant.

On remarque que les entreprises ne savent pas gérer correctement cet énorme flot de données, qui ne peut plus être restitué dans des délais acceptable. Il s'agit donc pour les principaux acteurs de trouver des moyens de les traiter et de les restituer au mieux dans un temps minimal.

De nouvelles solutions assez similaire ont donc été proposées, les acteurs proposent des appliances (solutions packagées alliant la partie software et la partie hardware) .En pratique, l'appliance représente une solution plus performante, plus maintenable et propose une réelle sécurité en terme de stabilité des performances de la solution.

Aujourd'hui, elles reposent toute sur la même technologie, appelée la technologie **« In-Memory ».**

A) La technologie « In-Memory »

1) Aspects techniques

Cette technologie repose principalement sur le stockage des informations en mémoire vive, et prend en compte les avancées technologiques au niveau matériel et logiciel.

En effet, la baisse significative des coûts de la mémoire vive permet la mise en place d'un système de stockage approprié aux données des grandes sociétés. L'accès rapide à ces données étant un point essentiel, un système de mise en cache très performant a été mis en place, prenant en compte la localité spatiale et temporelle des données traitées. Afin de faire face à la croissance du volume de données, des avancées dans la puissance de traitement sont nécessaires. C'est pourquoi, le système In-memory exploite le parallélisme des systèmes multiprocesseurs et des processeurs multi-cœur.

En plus des développements matériels, les avancées dans la technologie logicielle ont rendu le système « In-memory » possible. En effet, la mémoire vive a certes vu ses prix radicalement baissé, mais elle reste plus chère que le stockage en disque dur. Les techniques performantes de compression sont exigées pour gagner un bon compromis entre le coût de système et la performance. Ainsi, nous observons actuellement un facteur de compression de 5. Par exemple, avec 64Go de mémoire, le système peut stocker 320Go de données.

De plus, les bases de données peuvent-être gérées en colonne, selon les requêtes que l'on exécute.

2) Gestion des bases de données en colonne

Dans le cas d'une base de données orienté ligne, lors d'une requête, il est nécessaire de parcourir sur le disque toutes les colonnes des lignes traitées, afin de récupérer les champs correspondant aux colonnes nécessaire à la résolution de la requête.

Dans le cas d'un stockage en colonnes on travaille directement sur celles-ci, on va donc lire uniquement les colonnes qui sont nécessaires à la résolution de la requête. Comme toutes les lignes d'une colonne sont stockées de manière consécutive le temps de lecture de la colonne sera beaucoup plus rapide. Cela est particulièrement intéressant en BI car le nombre de colonnes lues pour répondre aux requêtes est faible (souvent moins de 5) et toujours bien inférieur au nombre de lignes.

Les colonnes peuvent aussi être triées. Ce qui réduit le nombre d'accès aléatoires dans le cas de recherches ce qui amène à une amélioration des performances. Mais il devient alors nécessaire de reconstruire entièrement la table en cas de modifications.

Exemple simplifié : Considérons cette base de données

EmpId	Lastname	Firstname	Salary
1	Smith	Joe	40000
2	Jones	Mary	50000
3	Johnson	Cathy	44000

Dans le cas d'une gestion en lignes, la base de données est stockée comme suit :

1,Smith,Joe,40000;
2,Jones,Mary,50000;
3,Johnson,Cathy,44000;

Nous avons donc bien un enregistrement par ligne.
Dans le cas d'une gestion en colonne, les enregistrements sont stockés de cette manière :

1,2,3;
Smith,Jones,Johnson;
Joe,Mary,Cathy;
40000,50000,44000;

Ce qui correspond à un enregistrement par colonne. Evidemment ceci n'est qu'une simplification et n'est pas représentatif de l'enregistrement au niveau physique, qui est considérablement modifié par l'indexation, le système de cache, etc.

Le stockage en colonne est donc meilleur que le stockage en ligne pour la recherche (donc particulièrement intéressant en décisionnel) et l'ajout, mais moins bon pour la modification (notamment le transactionnel).

3) Organisation et accès aux données

Le système « In-memory » a été non seulement conçu afin de garantir un accès rapide aux données, mais aussi afin de répondre aux exigences spécifiques des entreprises.

Ainsi, ce système permet l'exécution de procédures stockées, les algorithmes travaillent près des données, et soulagent ainsi la couche réseau. Une autre exigence concerne le vieillissement des données. En effet les sociétés stockent en moyenne des données pour une période de dix ans, mais utilisent que 20% des données les plus récentes. Ce système sépare donc en 2 les données : une partie passive et une partie active. La partie passive est en lecture seul, et peut-être stockée dans un matériel de stockage plus lent, en mémoire flash par exemple. La partie active est celle conservée dans la mémoire vive et utilisée par l'application. Cette dernière peut spécifier le comportement à adopter pour le vieillissement des données, le système effectue ensuite le passage de la partie active à la partie passive de la donnée de manière transparente.

La mise à jour des données est effectuée à l'aide d'une base temporaire stockant toutes les insertions et mises à jour effectuées par l'application. Ces mises à jour sont stockées sous forme d'insertions, afin de prendre en compte l'évolution des données dans le temps et de pouvoir gérer les accès concurrents. De temps en temps cette base est fusionnée avec la base courante.

Enfin, le système utilise sa propre gestion des threads : il organise les threads en 2 pools : un pour les requêtes analytiques, l'autre pour les requêtes transactionnelles. Une gestion automatique est assurée selon les requêtes utilisateurs.

Le système « In-Memory » propose donc, à terme, un unique système permettant de traiter les requêtes transactionnelles et analytiques[1].

B) Les Gains liés à cette technologie

1) L'occasion de mettre en relation des données différentes

Un premier gain directement lié à la vitesse est la réduction des temps d'exécution : Un traitement durant plusieurs nuits peut être terminé en 1h voir moins.

Si la capacité technologique à traiter des téraoctets de données à la très grande vitesse est acquise, cela n'augmente pas pour autant la valeur « métier » que les organisations peuvent en tirer : celle-ci dépend de la qualité et de la pertinence des analyses qui utilisent ces données. Ainsi, lorsque toutes les données sont disponibles, il s'agit de trouver les bonnes relations (ex : mettre en évidence les coûts de transaction et les coûts logistiques d'une livraison d'un produit, pour comprendre la rentabilité de chaque ligne de commande) avant de prendre les bonnes décisions.

2) Un reporting décloisonné

Aujourd'hui, la difficulté est de maintenir l'articulation entre les vues agrégées et les rapports détaillés dont ont besoin les différents niveaux opérationnels. Par exemple, il est compliqué de fournir à la direction générale des indicateurs agrégés tout en maintenant l'accès à des centaines de

[1] In-memory Data Management : An inflection point for Enterprise Applications Hasso Platner – Alexander Zeier

millions de lignes pour les opérationnels. Le drill-down est limité par la structure même des rapports et des data-marts, d'où des difficultés de compréhension et de coordination entre les différentes couches et périmètres organisationnels.

En permettant de manipuler à la volée de très gros volumes de données, la BI In-Memory élimine la nécessité technique de multiplier les niveaux d'agrégation.

Cette technologie commence à être disponible chez certains éditeurs. Son coût d'ensemble reste élevé car, en plus d'un hardware adapté et de nouvelles licences il faut que les ERP et les outils décisionnels soient sur des versions récentes. Des prototypes, appelés « Proof of Concept », permettent de bien mesurer les enjeux et le retour sur de tels investissements.

3) L'analyse prédictive

Un deuxième apport majeur de la technologie In-Memory est de rendre l'analyse prédictive beaucoup plus accessible en permettant d'utiliser des données de multiples domaines et des données externes. L'objectif est de croiser différentes sources pour analyser le passé et finir par des projections, qui servent de base au développement de stratégies d'action.

Au minimum on peut comparer dans le détail des transactions passées par des clients similaires et en déduire un plan d'action direct pour augmenter la contribution de l'un ou de l'autre. Un acteur de la grande distribution pourra beaucoup plus facilement croiser ses gigantesques volumes de données transactionnelles (des centaines de millions de tickets de caisse par exemple) avec, des interactions « sociales » générées à partir de plates-formes mobiles, des données météorologiques, de géolocalisation pour développer des stratégies opérationnelles à court ou moyen terme, ou encore des données comportementales générées à partir des réseaux sociaux[2].

C) Les différentes solutions

A partir de cette technologie, les grands éditeurs proposent de nouvelles solutions

- ***SAP***

SAP, leader mondial[3] a déjà mis sur le marché son application qui utilise cette technologie : HANA *High Performance Analytics Appliance :* une appliance matérielle et logicielle destinée à permettre aux applications compatibles de s'exécuter 'In-Memory'.

Chez les clients, elle fonctionne pour l'instant aux côté des systèmes existants. Les clients de la plateforme HANA font états d'un gain de temps de 1 000, 10 000, voire plus exceptionnellement 100 000[4].

Pour disposer de SAP HANA il faut pour l'entreprise, en plus de la couche matérielle, à minima des solutions SAP : l'ERP 6.0 avec EHP 4 ou 5, BW sur HANA[5].

[2] http://www.journaldunet.com/solutions/expert/50579/comment-la-bi--in-memory--decloisonne-la-prise-de-decision-et-accelere-la-capacite-d-action.shtml
[3] http://www.lemagit.fr/article/gartner-bi-donnees-analyse/10793/1/bi-analytique-sap-domine-marche-plus-2011/
[4] http://www.silicon.fr/hana-etat-des-lieux-de-loffre-in-memory-de-sap-1-72727.html

SAP est toujours en quêtes d'opportunités[6] liées au phénomène de Big Data, ainsi l'entreprise s'intéresse aux start-up, notamment en accueillant les plus innovantes et les plus intéressantes dans des forums qu'elle organise. Le but étant de réunir des entreprises pouvant être potentiellement rachetées. SAP a été intéressé par l'éditeur NextPrinciples[7], présent lors du 1er forum organisé par SAP propose un logiciel permettant de mettre en relation des données clients internes et celles en provenance des médias sociaux. Il s'agit ici de rajouter du contexte à l'information que l'on possède déjà.

Oracle

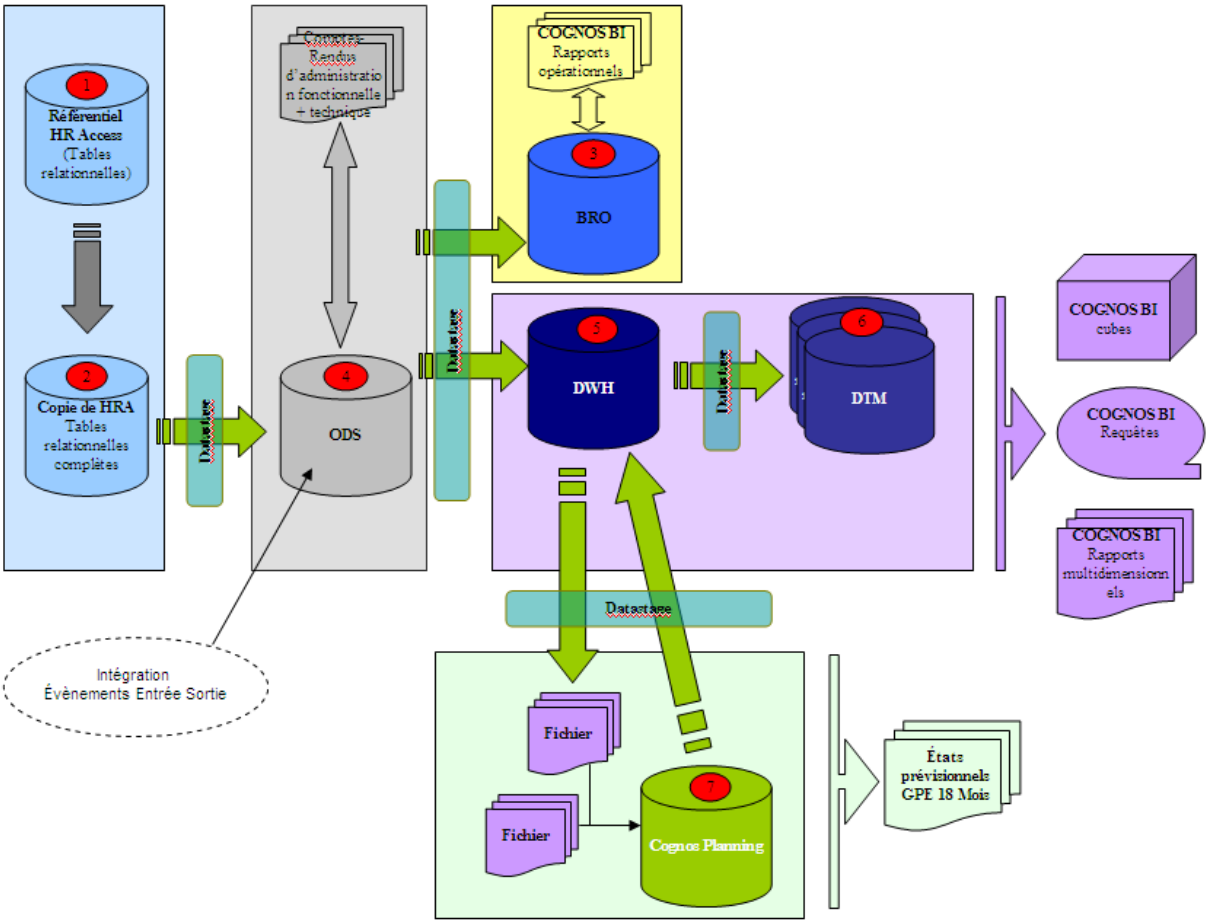
Oracle propose Exalytics[8], officialisé fin février qui repose sur une *appliance* composée d'un serveur Sun Fire, avec 1 To de RAM, 40 cœurs de processeurs Intel Xeon E7-4800, et une connectivité Infiniband 40 Gb/s et Ethernet 10 Gb/s. Côté applicatif, Exalytics embarque une version de la base de données TimesTen qui est la solution In-memory d'Oracle, avec une version mise à jour et optimisée de la suite analytique OBIEE (Business Intelligence Foundation Suite).

D'après Oracle, les tests comparatifs et les premiers retours des clients montrent que leur solution offre des gains de performance d'un facteur 10 à 100 pour le reporting OLAP relationnel (ROLAP) et les tableaux de bord, et d'un facteur pouvant atteindre 79 pour la modélisation OLAP multidimensionnelle (MOLAP).

[5] http://www.bilinksolutions.com/business-intelligence/2012/02/sap-hana-in-memory/
[6] http://www.decideo.fr/SAP-prepare-ses-prochaines-acquisitions-dans-le-Big-Data_a4975.html
[7] http://nextprinciples.com/
[8] https://blogs.oracle.com/CommunauteBI/entry/oracle_annonce_la_disponibilit%C3%A9_d

VI. Le datamining et l'alerte en temps réel (accès concurrents et performance)

L'objectif de cet axe était de trouver les suites logicielles actuellement sur le marché ayant trait au Data Mining et alertes temps réel, compatibles avec les couches stratégiques du système décisionnel de la SNCF, avec lesquelles elles seraient amenées à communiquer.



Les différentes couches stratégiques du système d'information décisionnel de la SNCF

On observe que c'est la solution Cognos de chez IBM qui est aujourd'hui déployée pour la couche stratégique "Restitution" et "Data Mining". C'est pourquoi nous avons privilégié nos recherches chez cet éditeur.

A) Real Time Monitoring sur Cognos 10

De plus en plus de domaines nécessitent de pouvoir travailler avec des données fraîches afin de prendre des décisions. Nous avons dans la partie (X) traitées des différentes solutions apportées en réponses aux défis provoqués par le stockage et l'exploitation en simultanée de ces données. Nous vous proposons dans cette section de découvrir quelles sont les suites logiciel actuellement proposées par Cognos 10 de chez IBM.

1) IBM Cognos Real-time Monitoring : introduction

IBM Cognos Real-time Monitoring est une solution de BI qui répond aux besoins de surveillance en temps réel d'une entreprise. Elle permet notamment la création de tableaux de bord interactifs, ainsi que le développement d'indicateurs clés de performance (ICPs).

Comme toute solution de surveillance en temps réel, CRTM permet l'accès à des informations stratégiques et donne ainsi possibilité de réagir rapidement à des opportunités d'économies. Elle délivre des fonctionnalités décisionnelles telles que la surveillance en temps réel d'ICPs et de mesures, sensibles au temps, ainsi que des outils permettant d'anticiper les situations de crise grâce à un système d'alerte proactif. Les responsables ainsi que les analystes ont un aperçu immédiat des divers changements concernant l'environnement opérationnel de l'organisation.

2) Accès rapide aux données

IBM Cognos Real-time Monitoring utilise un serveur analytique in-memory de 64-bit afin de capturer les données temps réels alimentant les tableaux de bords, les rapports utilisateurs etc. Grâce à cette technologie in-memory décrite dans la section précédente, les utilisateurs peuvent rapidement accéder en temps réel à l'information provenant de divers systèmes opérationnels sources pour surveiller, analyser et effectuer ces rapports. Les serveurs in-memory profitent de technologies de stockage de pointes pour une intégration continue des données, une sécurité optimale, des analyses multidimensionnelles, des modélisations dynamiques, l'exécution des règles de gestion, ainsi que le traitement des exceptions et des rapports d'alertes.

Il est intéressant de remarquer que la solution délivre un temps de réponse optimal, même si beaucoup d'utilisateur la sollicite grâce à une architecture faite pour absorber la monté en charge.

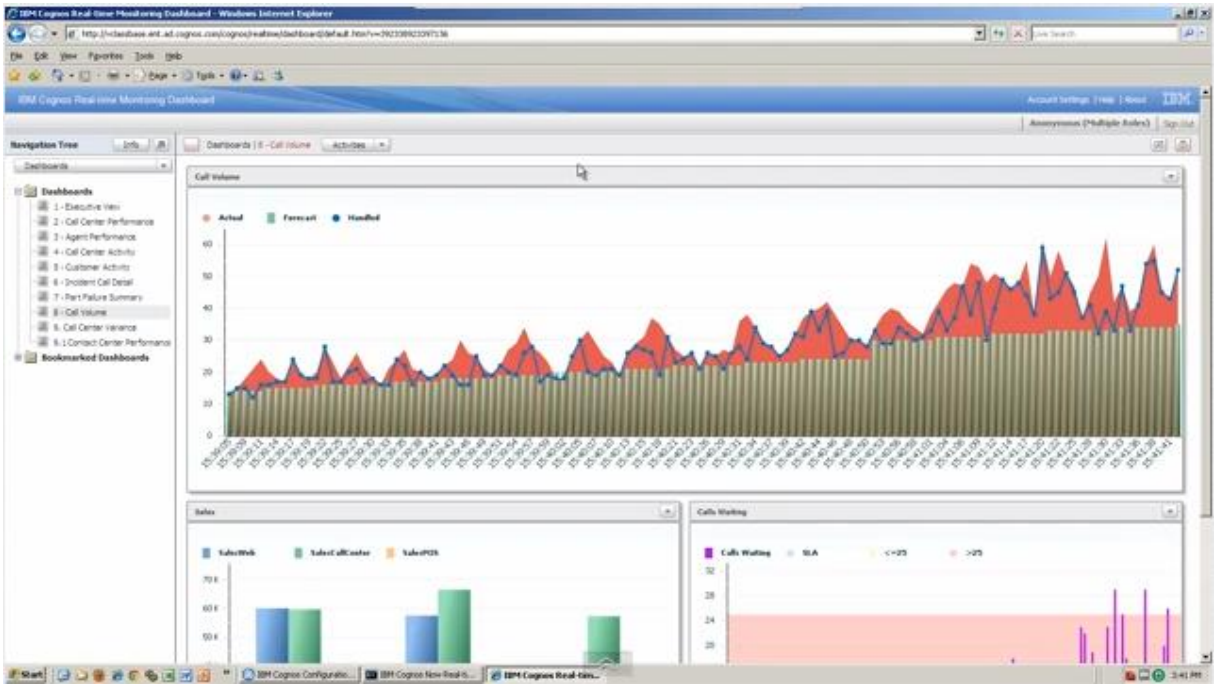


Figure 4 Vue des indicateurs clés de performance ainsi que des données opérationnelles

Conclusion

Au vu de l'architecture actuelle du système d'information décisionnel de la SNCF, la solution Cognos de surveillance en temps réel nous paraît être la plus indiquée pour s'y insérer en douceur. Elle permet de fusionner en un unique outil les deux axes de veille du Data Mining et de l'alerte en temps réel. Il semble de plus que les technologies de stockage in-memory dernières générations, soient une excellente réponse aux problématiques d'exploitation de données fraîches abordées dans la partie 2-V de ce livre blanc.

VII. Les législations concernant la Business Intelligence Temps réel

La Business Intelligence temps réel est une technologie se basant sur des informations issues de données stockées. Ces données peuvent regrouper des informations libres mais bien souvent dans le décisionnel, ce sont les clients, et par extension leurs données personnelles, qui sont au cœur de cette technologie.

Les bases de données sont soumises à des réglementations mais aucune réglementation universelle n'est, à ce jour, existante.

D'autre part, nous voyons que les principaux acteurs de la BI temps réel sont les opérateurs en bourses et notamment la HFT (High Frequency Trading). Cette technique est soumise à controverse car jugée responsable de la crise économique de 2008. C'est pourquoi de nombreuses organisations et gouvernement sont actuellement en train d'affiner des lois pour limiter cette technique, voire même la supprimer.

A) Les législations locales

Les bases de données sont de plus en plus réparties sur l'ensemble du globe. C'est pourquoi le cadre juridique est assez instable. La principale question est de savoir quelle législation est en vigueur car ces dernières sont différentes entre chaque pays.

Actuellement, c'est la localisation des données qui influence le type de legislation en vigueur. Par exemple si une entreprise basée à Lyon stocke ses données à Pékin, alors ces informations seront soumises aux lois en vigueur dans le pays.

B) La législation française

En France, certaine lois concernant les données personnelles sont soumises à des restrictions. C'est la CNIL (Commission Nationale de l'Informatique et des libertés) qui est en charge de faire respecter les normes françaises dans les pays.

1) La sécurité des fichiers

Tout responsable de traitement informatique de données personnelles **doit adopter des mesures de sécurité physiques** (sécurité des locaux), **logiques** (sécurité des systèmes d'information) et **adaptées** à la nature des données et aux risques présentés par le traitement.

Le non-respect de l'obligation de sécurité est sanctionné de 5 ans d'emprisonnement et de 300 000 € d'amende.
art. 226-17 du code pénal

2) La confidentialité des données

Seules les personnes autorisées peuvent accéder aux données personnelles contenues dans un fichier. Il s'agit **des destinataires** explicitement désignés pour en obtenir régulièrement communication et **des «tiers autorisés»** ayant qualité pour les recevoir de façon ponctuelle et motivée (ex. : la police, le fisc).

La communication d'informations à des personnes non-autorisées est punie de 5 ans d'emprisonnement et de 300 000 € d'amende.

La divulgation d'informations commise par imprudence ou négligence est punie de 3 ans d'emprisonnement et de 100 000 € d'amende. art. 226-22 du code pénal

3) La durée de conservation des informations

Les données personnelles **ont une date de péremption**.

Le responsable d'un fichier fixe **une durée de conservation raisonnable** en fonction de l'objectif du fichier.

Le code pénal sanctionne la conservation des données pour une durée supérieure à celle qui a été déclarée de 5 ans d'emprisonnement et de 300 000 € d'amende.

art. 226-20 du code pénal

4) L'information des personnes

Le responsable d'un fichier doit permettre aux personnes concernées par des informations qu'il détient d'exercer pleinement leurs droits. Pour cela, il doit leur communiquer : son identité, la finalité de son traitement, le caractère obligatoire ou facultatif des réponses, les destinataires des informations, l'existence de droits, les transmissions envisagées.

Le refus ou l'entrave au bon exercice des droits des personnes est puni de 1500 € par infraction constatée et 3 000 € en cas de récidive.
art. 131-13 du code pénal ***Décret n° 2005-1309 du 20 octobre 2005***

5) L'autorisation de la CNIL

Les traitements informatiques de données personnelles qui présentent des risques particuliers d'atteinte aux droits et aux libertés doivent, **avant leur mise en oeuvre**, être soumis à l'autorisation de la CNIL.

Le non-accomplissement des formalités auprès de la CNIL est sanctionné de 5 ans d'emprisonnement et 300 000€ d'amende. art. 226-16 du code pénal

6) La finalité des traitements

Un fichier doit avoir **un objectif précis**.
Les informations exploitées dans un fichier doivent être **cohérentes par rapport à son objectif**.
Les informations **ne peuvent pas être réutilisées de manière incompatible avec la finalité** pour laquelle elles ont été collectées.

Tout détournement de finalité est passible de 5 ans d'emprisonnement et de 300 000 € d'amende.art. 226.21 du code pénal

7) Le transfert des données

Le transfert de données depuis la France vers un pays en dehors de l'Union Européenne est soumis à des restrictions. En France, ces transferts sont interdits en dehors des motifs suivants :

- Si le transfert a lieu vers un pays reconnu comme "adéquat" par la Commission européenne. C'est le cas du Canada, de la Suisse, de l'Argentine, des territoires de Guernesey, de Jersey et de l'Isle de Man
- Si des Clauses Contractuelles Types, approuvées par la Commission européenne, sont signées entre deux entreprises
- Si des Règles internes d'entreprises (BCR) sont adoptées au sein d'un groupe
- Si dans le cas d'un transfert vers les États-Unis, l'entreprise destinataire a adhéré au Safe Harbor,
- Si l'une des exceptions prévues par l'article 69 de la loi Informatique et Libertés est invoquée

Les sanctions encourues en cas de non respect des règles en matière de transferts sont de 300 000 euros d'amende et de 5 ans d'emprisonnement. (Articles 226-16, 226-16 A et 226-22-1 du Code pénal)

C) L'utilisation des données

La CNIL réglemente l'utilisation de données personnelles. Mais si les conditions d'utilisations (préalablement approuvées par cet organisme) ont été accepté par l'utilisateur lors de son inscription, alors l'entreprise peut utiliser ces données (dans le respect de charte).

a. Les réseaux sociaux

Les réseaux sociaux sont actuellement la plus grande source d'information à propos des clients. De nombreuses entreprises s'intéressent donc à ces données et au vu de la grande quantité d'information, se base sur la BI TR. Mais quand est-il de l'accès à ces données ?

Facebook lui utilise une politique de confidentialité soumise à contestation, mais qui regroupe toute les utilisations des données que celui-ci peut faire ainsi que la politique de l'accès par des entreprises ou des sites internet extérieur à ces données.

Nous pouvons donc voir que les fichiers de données personnelles sont conservés pendant 180 jours et les données de géo localisation ne sont existantes que pendant la durée de la connexion.

Des données peuvent être échangées avec les partenaires de Facebook, mais l'identité de la personne ne sera pas communiquée.

L'accès aux « post » n'est possible qu'après autorisation de la personne concernée ou suivant la visibilité de son post (public ou non).

Twitter lui utilise des données de plusieurs sortes : les données de connexion qui sont stocké pendant une durée de 18 mois, et la mémorisation de la navigation. Twitter s'engage à ne pas divulguer les informations personnelles sans un consentement préalable de la personne. L'accès au « tweet » est quant-à-lui accessible par n'importe quelle personne.

b. Les données mobiles

Les données mobiles sont des informations dites d'itinérance qui permette de garantir une meilleure couverture réseau de votre appareil mobile. Ces données sont utilisées par les publicitaire afin d'analyser les flux de population.

Ces informations sont stockées dans les serveurs de la téléphonie mobile et ne peuvent pas être utilisée par une entreprise tierce. Sauf si ces informations sont transmises sans identification de la personne.

Les OS des smartphones utilisent également les données de géo-localisation afin de permettre un accès à des services proches de la position actuelle.

Les deux géants actuels concernant l'utilisation des données mobiles (Apple et Google) utilisent une politique assez similaire, en utilisant les données qu'après l'accord de l'utilisateur et/ou rendant anonymes les informations personnelles.

D) La législation des opérations sur la bourse

Montré du doigt depuis 2008, les prises de décisions influant sur la bourse sont actuellement l'une des industries majeures utilisant la BI temps réel. Ce sont surtout les transactions à haute fréquence, qui sont considéré comme responsable de la crise économique de 2008. En effet, pendant les quelques minutes précédent la crise, l'équivalent de cinq fois l'économie mondiale qui s'est échangée en quelques minutes.

De nombreuses organisation non gouvernementale ainsi que l'AMF (l'Autorité des marchés financiers), dénonce cet abus de pouvoir, ainsi une première requête d'interdiction de cette pratique (et ainsi de la principale forme de BI temps réel), qui a été rejetée par le gouvernement français en 2011.

Une directive européenne est en cours d'adoption depuis le 8 mars 2012 et vise à limiter l'utilisation de la BI temps réel dans les marchés financier. Voici les points de restrictions :

- Les entreprises ayant recours au HFT se verront ainsi dans l'obligation de mettre en place des procédures de contrôle des risques avancées pour prévenir tout dysfonctionnement potentiel des algorithmes utilisés. Le descriptif de ces algorithmes devra par ailleurs être mis à disposition du régulateur.
- Les high frequency traders exécutant un nombre significatif d'ordres sur un instrument particulier devront ensuite être en mesure de fournir en continu un niveau de liquidité minimum sur cet instrument (comme sont contraints de le faire les marketmakers actuellement)
- Les high frequency traders devront par ailleurs observer un temps de latence minimum avant d'annuler les ordres émis
- Un ratio minimum d'exécution des ordres passés dans le cadre du HFT pourrait enfin être exigé ce qui passerait par la mise en place de pénalités financières pour les ordres annulés.

Nous pouvons donc observer que la restriction du temps de calcul, et l'imposition d'un temps minimal a été exclu, et donc ces réglementations ne pourront pas mettre en péril l'utilisation de la BI temps réel au sein de ses clients principaux.

VIII. BI mobile

La BI mobile prend de l'ampleur à mesure que les appareils se développent. L'invasion du marché par les tablettes et Smartphones associé à la requête grandissante des utilisateurs en BI pseudo temps réel ont donné lieu à la BI mobile.

En effet les utilisateurs souhaitent davantage de réactivité aux données. La BI mobile permet la prise de décision instantanée et l'alerte en pseudo temps réel. L'application Androïde de BPM conseil supporte la lecture de code barre et « *Dans le secteur hospitalier, notre application est utilisée sur tablette Androide : en barcodant un médicament dans la chambre depuis la tablette, on peut accéder aux stocks en pharmacie, en Wifi. De la même façon, dans la grande distribution, l'application permet des relevés de prix directement dans les linaires* ».

SAS, SAP, IBM, tous les éditeurs de solutions BI ont développé des outils de BI mobile. Parmi les éditeurs à s'être rapidement positionnés, citons Business Objects, Cognos et Oracle. Les utilisateurs ont suivi la vague du Bring Your Own Device (apportez votre propre appareil) et se sont munis d'Ipad, Iphone et autres appareils Androïdes. Les éditeurs se retrouvent maintenant confrontés à un choix : développer leurs applications en HTML ou créer de nouvelles applications.

Microstrategy a choisi de miser sur des applications Iphone et Ipad alors que Jaspersoft, QlikTech et Prelytisse ont optés pour le HTML. En effet le HTML semble permettre un meilleur affranchissement par rapport au terminal et permet de proposer différentes interfaces en fonction de l'appareil utilisé. BPM conseil a quant à lui lancé une application Androïde en 2010 afin d'accéder aux données de sa plateforme décisionnelle Vanilla.

Selon une enquête conduite par TWDI début 2012, l'Ipad et l'Iphone reste les deux choix favoris des utilisateurs de la BI mobile. De même les outils collaboratifs et les tableurs restent les deux premiers cas d'utilisation d'un appareil dans le cadre de la BI mobile.

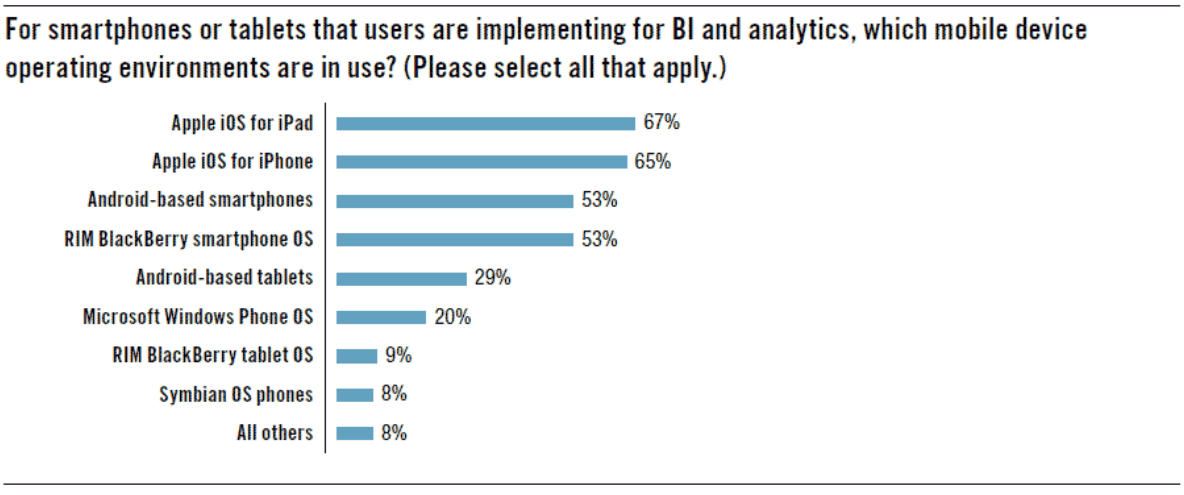


Figure 5 - Les appareils choisis pour la BI mobile

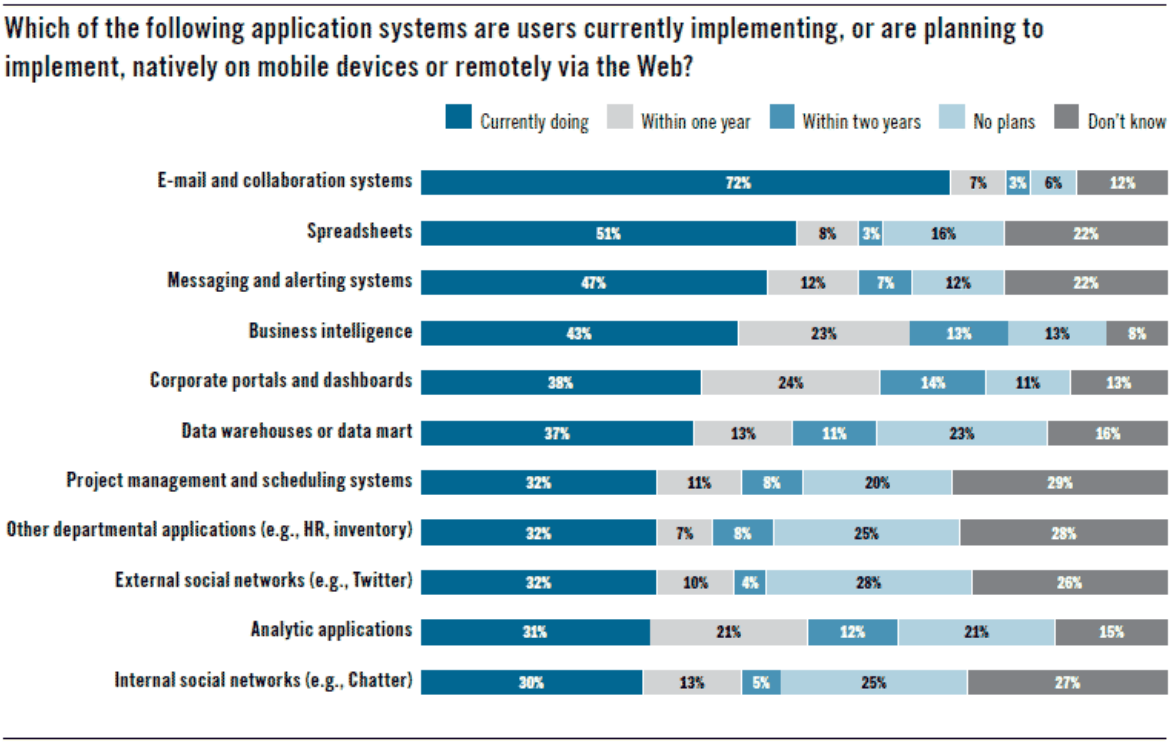


Figure 6 - Quelle utilisation des appareils ?

Cependant, bien que la BI mobile se développe, elle reste utilisée par une petite partie des employés. La plupart des utilisateurs sont les managers et responsables de l'entreprise et la totalité des utilisateurs représente en moyenne 10% des effectifs de l'entreprise.

Partie 3 : l'avenir de la Business Intelligence temps réel

I. Scrutation des réseaux sociaux

L'une des idées de l'utilisation de la BI Temps Réel, est de l'utiliser afin de capter dans un laps de temps très court l'avis des clients de la SNCF, ce qui permettrait de détecter rapidement les anomalies qui ont pu survenir et ainsi répondre au mieux aux besoins des utilisateurs.

L'idée est d'utiliser les bases de données des réseaux sociaux connus et de capter le mécontentement des utilisateurs en vu de les informer d'une éventuelle panne, ou d'une réaction non prévue sur une des lignes. Il serait alors également possible d'informer en temps réel aux utilisateurs des problèmes que rencontre le train.

II. Outil de gestion des correspondances

Ce système se baserait sur les bases de données temps réel des transports de correspondance (taxi, transport en communs, avions ou encore ferrys) afin de faire profiter en temps réel du meilleur acheminement possible aux clients utilisant ce service.

III. Analyse des flux de personnes pour analyser sur le trafique

De plus en plus de personnes possèdent actuellement des smartphones. Ces appareils envoient constamment des données de positionnement à des antennes relais. L'analyse de ces données permet actuellement aux publicitaires de déterminer où se situent les flux de personnes afin d'afficher une affiche publicitaire dans des lieux stratégiques.

La SNCF pourrait utiliser ces données pour analyser où se situent les flux de personnes, où ils prennent place, et ainsi prévoir par du datamining où pourraient avoir lieu les pics d'activité de flux de personnes et ainsi répondre au mieux aux besoins des clients.

Ces données peuvent également servir à déterminer les itinéraires non existants et ainsi développer le réseau ferroviaire.

IV. Une application smartphone

Dans le but d'améliorer continuellement la satisfaction de ses clients, la SNCF pourrait mettre en place une application smartphone permettant de conseiller un passager dans le cas d'un imprévu lié à son voyage (correspondance annulée, retardée etc.). Un QR code sur le billet serait scanné par le smartphone, et après analyse du trajet et des problèmes associés à celui-ci (s'il y en a), le système propose au client plusieurs alternatives, à savoir patienter si le retard est de l'ordre du quart d'heure, offrir une boisson à la gare de correspondance si le retard excède la demi-heure, proposer une sélection d'hôtels avec réduction si annulation de la correspondance etc... Avec la possibilité pour le passager de donner son avis (positif, négatif) sur la proposition qu'il a choisi afin que le système apprenne avec le temps à prévoir ce dont le client aura envie/besoin selon une situation donnée. Beaucoup de paramètres peuvent être pris en compte pour suggérer au client une alternative (la température extérieure, l'heure, le temps d'attente, les partenariats avec les commerçants autours de la gare, le sexe du client, son programme de fidélité etc.).

Bibliographie

REFERENCES AXE CLOUD BI :

http://mysaas.fr/

http://csrc.nist.gov/

http://www.journaldunet.com/solutions/expert/49968/le-cloud-et-la-business-intelligence--une-tendance-de-fond-pour-les-entreprises.shtml

http://blog.agiledss.com/fr/bid/81279/Le-Cloud-BI-le-point-sur-le-prochain-bouleversement-technologique-Partie-1-de-2

http://tdwi.org/articles/2012/01/24/2012-year-of-cloud-bi.aspx

http://tdwi.org/Articles/2012/03/13/Cloud-BI-Progress-and-Pitfalls.aspx

http://dsi.silicon.fr/competences-rh/bi-dans-le-cloud-distinguer-fiction-et-realite-1806

http://www.slideshare.net/businessintelligenze/businessintelligenze-on-cloud-bi-english-7515247

http://www.legrandbi.com/

http://www.europe1.fr/Dossiers/Les-cles-du-cloud/Articles/Cloud-et-securite-des-donnees-la-premiere-preoccupation-des-entreprises-799287/

http://www.sap.com/

http://www.scribd.com/doc/67865933/SAP-Business-Objects-BI-on-Demand

http://www.silicon.fr/sap-et-success-factors-la-roadmap-de-loffre-rh-privilegie-le-cloud-72129.html

http://www.oracle.com/us/solutions/business-intelligence/cloud-ready-oracle-bi-177505.pdf

http://www.oracle.com/us/solutions/cloud/overview/index.html

http://www.jaspersoft.com/fr/d%C3%A9cisionnel-et-cloud

http://www.bittle-solutions.com/

http://fr.bimeanalytics.com/blog/we-are-cloud-launches-first-genuine-cloud-bi-service-in-china-press-release/

REFERENCES AXE LEGISLATION :

www.cnil.fr/

http://www.facebook.com/about/privacy/

https://twitter.com/privacy

http://finance.sia-conseil.com/20120221/le-trading-haute-frequence-n%E2%80%99echappera-pas-a-la-regulation/

http://www.euractiv.fr/trading-haute-frequence-parlement-serre-vis-article

http://www.bfmbusiness.com/toute-linfo-eco/fiscalit%C3%A9-politique-fin-publiques/le-trading-haute-fr%C3%A9quence-dans-le-viseur-du-gouve

http://www.abcbourse.com/apprendre/18_le_trading_haute_frequence.html

http://www.amf-france.org/

REFERENCES AXE BI MOBILE :

http://www.legrandbi.com/2011/04/bi-mobile/

http://pro.01net.com/editorial/531241/dossier-bi-mobile-le-decisionnel-va-sur-le-terrain/

http://www.legrandbi.com/2012/01/bi-mobile-utilisateurs/

http://www.infoworld.com/d/business-intelligence/sas-pushes-bi-the-ipad-and-iphones-846