

Le Big Data & l'Intelligence Economique

Tout comme pour l'Intelligence Economique, la matière première du Big Data est l'information, et les similitudes entre les deux concepts ne s'arrêtent pas là. Le Big Data et l'Intelligence Économique partagent les grandes phases du traitement de l'information : Recherche, Collecte, Analyse et Diffusion. Basé sur de nombreuses technologies nécessitant des compétences et apports techniques importants, le Big Data complexifie autant qu'il simplifie le domaine de la gestion de l'information. Efficacement exploitée, l'importante masse de données que peuvent traiter les technologies du Big Data est un atout non négligeable pour le professionnel de l'Intelligence Economique. Se voulant la plus exhaustive possible, cette approche du traitement de l'information à grande échelle permet effectivement de confronter un maximum de paramètres et d'éléments contextuels pour une aide à la prise de décision la plus juste et objective.

NOTRE DEFINITION

« Nous créons actuellement en deux jours autant d'information que nous en avons créée depuis la naissance de la civilisation jusqu'en 2003 », rappelait récemment Eric Schmidt, Chairman de Google. Pour les entreprises, le défi consiste à traiter et activer les données disponibles afin d'améliorer leur compétitivité. Aux données « classiques » déjà manipulées par les entreprises et exploitées par les techniques de Business Intelligence, s'ajoutent désormais les données informelles, essentiellement issues du « crowdsourcing », via les médias sociaux, les terminaux mobiles, et, de plus en plus via les senseurs intégrés aux objets.

Nous choisissons ici de décomposer le Big Data en trois composants, le premier étant la collecte rapide et le stockage d'un ensemble de données volumineuses et variées, le second l'analyse et le traitement de ces données, et le dernier élément étant la diffusion par la visualisation et la compréhension de ces données. Le Big Data oblige les entreprises à redéfinir non seulement leurs processus et technologies mais également leurs métiers. Les perspectives et opportunités liées au traitement des Big Data, c'est-à-dire aux « analytics », semblent sans limites et concernent des domaines aussi variés que les tendances industrielles, l'épidémiologie, le marketing ou la sécurité. Le phénomène du Big Data est unanimement considéré comme l'un des grands défis informatiques de la prochaine décennie.

Pourquoi Big et pourquoi maintenant ?

Trois facteurs expliquent le développement du Big Data :

1. **Le coût du stockage.** Celui-ci ne cesse de diminuer et est un paramètre de moins en moins déterminant pour les entreprises. Par exemple, les solutions de Cloud Computing permettent une gestion élastique des données, s'adaptant aux besoins en temps réel des entreprises.
2. **Les plateformes de stockage distribuées et les réseaux à très haut débit (THD).** Avec le développement du THD et l'avènement du Cloud Computing, le lieu de

stockage des données n'est plus un critère primordial. Celles-ci sont désormais stockées à des endroits physiques distincts, et parfois non identifiés. Ainsi, quel utilisateur sait où se trouvent réellement ses photos postées sur Facebook ou ses messages sur Google Mail ?

3. **Le développement des nouvelles technologies de gestion et d'analyse de données.** Parmi ces solutions technologiques liées au Big Data, l'une des références est la plateforme Hadoop (Apache Foundation) permettant le développement et la gestion d'applications distribuées adressant des quantités de données abondantes et évolutives.

Ces trois facteurs combinés tendent à transformer la gestion et le stockage des données en un « simple » service.

Sources et types de données

Afin de comprendre le phénomène Big Data de façon concrète, il est intéressant d'identifier les sources de production des données :

- Les applications et services professionnels (ERP, CRM, etc.)
- Le Web
- Les médias sociaux
- Le mobile
- Les objets

Il existe deux grandes catégories de données:

- **Les données structurées.** Ce sont les données que l'on peut facilement organiser par rangées et colonnes, et qui sont traditionnellement gérées dans des bases de données. Il s'agit notamment des données liées au fonctionnement habituel des entreprises et organisations (stocks, comptabilité, finances, ressources humaines, statistiques, études scientifiques, etc.).
- **Les données non structurées.** Déjà présentes sous la forme de la production bureautique non organisée, ces données se multiplient de manière exponentielle et incontrôlable avec les plateformes de crowdsourcing, le mobile et l'Internet des objets. Par leur volume, leur vitesse d'acquisition, et la variété de leurs formats, elles nécessitent de nouveaux outils pour leur stockage, leur traitement et leur analyse. C'est à leur développement que correspond la naissance du Big Data.

Les 5 V du Big Data

Volume

Le volume est la première caractéristique qui vient à l'esprit lorsqu'on aborde le Big Data. Suivant une courbe exponentielle, ce volume concerne non seulement les données produites chaque jour, mais aussi celui des capacités de stockage des supports informatiques. Il tend aujourd'hui vers l'infini et nécessite des unités de mesure comme le yottabyte qui équivaut à 1.000.000.000.000.000.000.000.000 de bytes, soit un trillion de terabytes. Les technologies du Big Data interviennent alors : les bases de données et outils de gestion traditionnels ne sont en effet pas en mesure de gérer de telles quantités de données.

Vitesse

La vitesse du Big Data représente le temps nécessaire pour que les données soient collectées, traitées et activées par l'entreprise. Le monde digital est désormais « plus rapide que le temps réel » et la gestion de données n'échappe pas à cette tendance. Celles-ci sont produites, collectées, traitées, et partagées à une vitesse inédite. Comme l'a justement remarqué Rupert Murdoch : « The world is changing very fast. Big will not beat small anymore. It will be the fast beating the slow ».

Une entreprise de services financiers doit ainsi traiter et analyser plusieurs millions de messages par seconde pour activer ou non des ordres sur les marchés. Le temps réel est la nouvelle unité de temps pour les entreprises et les systèmes classiques de traitement des informations (gestion, personnalisation, marketing, etc.) se révèlent, là encore, peu adaptés.

Variété

La montée en puissance des données non structurées s'accompagne d'une diversification des formats et des types de données. L'entreprise doit par exemple donner du sens aux avis et propositions émis sur Facebook, aux images, aux sons, aux vidéos, mais aussi aux informations émises par les terminaux mobiles ou issues des interactions M2M (Machine To Machine). Dans le domaine du commerce, des solutions Big Data permettent de relier les données non structurées émises par un client (comportement, intonations de la voix, etc.) aux données classiques enregistrées à son sujet (historique des achats, service après-vente, etc.) pour développer en temps réel une offre adaptée à ses besoins.

L'Américain TerraEchos a ainsi développé un système très sophistiqué pour classer les sons. Sa technologie permet un contrôle intelligent et en direct pour un périmètre donné, équipé de milliers de capteurs. Ceux-ci recueillent les sons détectés qui sont ensuite organisés et analysés pour déclencher des actions appropriées (envoi d'un garde ou activation d'une caméra par exemple) sur la base de modèles à la fois prédéfinis et évolutifs.

Véracité

La qualité et la fiabilité des données est un paramètre essentiel. Cette information est d'autant plus vraie dans un contexte où les sources de données sont désormais majoritairement hors du périmètre de contrôle des organisations. Le concept de véracité traduit donc le besoin stratégique de disposer de données de qualité.

En principe, une plateforme Big Data permet à une entreprise d'analyser les données relatives à son environnement de manière quasi exhaustive et donc d'améliorer sa compréhension de l'ensemble des composants de son environnement (clients, partenaires, produits, concurrents, ...). Paradoxalement, l'entreprise est confrontée au risque de se noyer dans cet océan de données et de se trouver dans l'incapacité de faire le tri entre les informations pertinentes et le « bruit ».

Une bonne illustration de ce problème est survenue lors de l'élection présidentielle de 2012 au Mexique, quand des tweets issus de robots spammeurs et des faux comptes « orientés » ont largement pollué le débat politique et son analyse sur Twitter.

Valeur

En bout de course, la valeur du Big Data pour une entreprise se mesurera à l'avantage compétitif qu'elle en aura dégagé. Cela dépend notamment de la qualité des analytics et

de la compétence des « data scientists » chargés de leur donner du sens. Créer des données pour le plaisir de la performance technique n'est pas viable sur le long terme. Ces nouvelles activités impliquent également des outils de mesure du ROI.

Data visualisation

Comme son nom l'indique, la data visualisation est la visualisation des données du big data. Une visualisation intelligente et opérationnelle des analytics est un facteur clé dans le déploiement du Big Data dans les entreprises. Le développement des infographies va d'ailleurs de pair avec le développement des techniques d'exploitation des données.

La datavisualisation permet de :

- **Rendre les données directement compréhensibles.** Tandis que les tableaux de données classiques deviennent difficilement lisibles, les diagrammes, graphiques ou cartes, permettent une compréhension rapide et aisée des données ;
- **Révéler les détails.** La visualisation des données exploite la capacité de la vue humaine à envisager une image dans son ensemble, tout en captant différents détails qui seraient passés inaperçus dans un format textuel ou dans une feuille de calcul ;
- **Fournir des réponses rapides.** En éliminant le processus de requête, la visualisation de données réduit le temps nécessaire pour dégager des informations pertinentes pour les entreprises, par exemple au sujet de la fréquentation d'un site Web ;
- **Prendre de meilleures décisions.** En permettant de visualiser les modèles, les tendances et les relations issues de l'analyse des données, l'entreprise peut améliorer la qualité de ses processus de prise de décision ;
- **Simplifier les analyses.** Les datavisualisations doivent être interactives. Les outils Webmasters de Google en sont l'illustration. En offrant des fonctionnalités simples et instinctives pour modifier les jeux de données et les critères d'analyse, ces outils libèrent la créativité des utilisateurs.